Noise in Speech Signal Processing

Noise

Noise causes trouble in SSP!

- Interference in phone lines, recordings.
- Offsets in ASR.

So, we want to compensate for it.
To compensate, we need to know how it behaves.

Convolutional noise

Convolutional noise is really an (unwanted) filter.

- The effect of the air between person and microphone.
- The effect of a telephone channel filter.
- A remnant of an imperfect microphone (all microphones?)

Convolutional “noise” is not necessarily noise

- Unless we don’t actually want it there.
- ...which is most of the time.

It is the driving force behind the homomorphic signal processing framework of Oppenheim.

Recap

Convolution in homomorphic processing.

$$x = s * h$$

After DFT,

$$X = S \times H.$$

And then after $\log$,

$$\log X = \log S + \log H.$$

Summary

Convolutional noise is a problem that must be addressed.
However, generally speaking, it can be handled in the cepstral domain.
This section will say nothing more about convolutional noise.

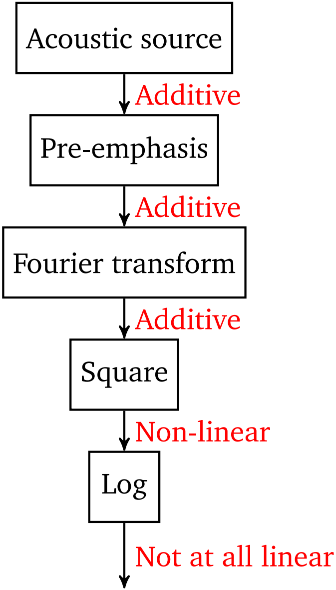
Additive noise

Additive noise is

- Noise that is added to the signal (Duh...)
- Other things going on near a microphone.
  - Traffic.
  - Fans.
- "Static" or radio interference added in the channel.

It may or may not be intrusive.

The homomorphic approach to additive noise



MFCC processing

- The first few stages are linear.
- It becomes non-linear after the square.
- Subsequently becomes very non-linear.

We can deal with additive noise in a linear fashion either in time or frequency domain, but not power domain.

Frequency domain processing

Noise has a spectrum

- Typically not flat.
- Close to “pink”.
- Typically biassed towards low frequencies.

Speech also has a spectrum

- Biassed towards $\approx$1kHz.
- Little power $<$ 300Hz.

Time domain processing would not take account of this.

Noise as a random variable

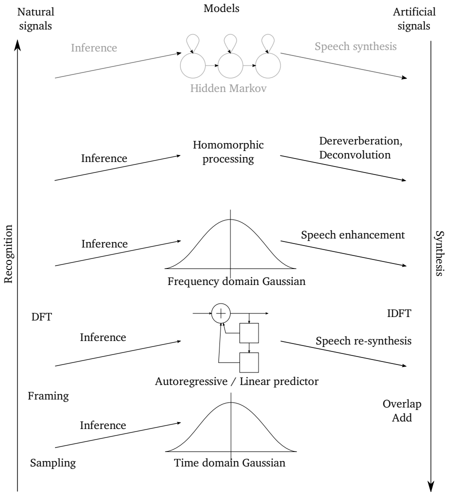
The nature of noise is random

- We have to treat it as a random variable, not as a deterministic signal.
- We have to understand how the random variable behaves through a DFT.

The following sections address how to deal with noise in frequency domain.

The Gaussian model

A map of speech signal processing



The Gaussian assumption

It's not a great leap of faith to assume noise is Gaussian

- Rather, individual DFT bins are (complex) Gaussian.
- The noise has a spectral shape.

It is possible to assume that speech is also Gaussian

- Clearly it isn't! Not long term anyway.
- Within a phone, however, it's closer to Gaussian.

How good an assumption

Fricatives and the like are basically noise

- It's safe to assume Gaussian

Vowels are not really noise

- Perhaps they are for a (whispered) noise excitation.
- Typically, though, they're vocalised.

In the vocalised case, the noise assumption is more an assumption about the variation of the vocalisation.

- It is healthy to retain some skepticism!

Ultimately, it's better to proceed with some measure of uncertainty rather than none at all.

Sum of two Gaussian RVs

If both speech and noise are Gaussian, we are concerned with the sum of two Gaussian distributions.
If we have two (zero mean) Gaussian RVs, s and n, their joint PDF is

$$p(s, n) = \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left(-\frac{s^2}{2{\sigma_s}^2}\right) \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{n^2}{2{\sigma_n}^2}\right).$$

However, we do not observe s and n together. We actually observe $t = s + n$.

Write as one variable and sum

Making the change of variable,

$$t = s + n, \qquad s' = s,$$

the Jacobian determinant is

$$J(s', t) = \begin{vmatrix} \dfrac{\partial s}{\partial s'} & \dfrac{\partial n}{\partial s'} \\ \\ \dfrac{\partial s}{\partial t} & \dfrac{\partial n}{\partial t} \end{vmatrix} = 1,$$

giving

$$p\left(s, t\right) = \frac{1}{2\pi\sigma_s\sigma_n} \exp\left(-\frac{s^2}{2{\sigma_s}^2} - \frac{(t-s)^2}{2{\sigma_n}^2}\right).$$

Integrate

To get the distribution of the sum, integrate over s

$$p(t) = \int_{-\infty}^{\infty} ds \, \frac{1}{2\pi\sigma_s\sigma_n} \exp\left(-\frac{s^2}{2{\sigma_s}^2} - \frac{(t-s)^2}{2{\sigma_n}^2}\right).$$

This is a convolution of two Gaussians. The result is very standard.

- Albeit messy. Proved in appendix in the notes.

$$p(t) = \frac{1}{\sqrt{2\pi({\sigma_s}^2 + {\sigma_n}^2)}} \exp\left(-\frac{t^2}{2({\sigma_s}^2 + {\sigma_n}^2)}\right).$$

Variances add

This is a very general rule:

> *The sum of two Gaussian random variables is another Gaussian RV with variance equal to the the sum of the individual variances.*

Or even more generally:

> *A linear transform of a vector of Gaussian RVs is another vector of Gaussian RVs.*

Corrollary:

- The DFT of white noise is white noise.

The Gaussian assumption revisited

We can justify the assumption from two different directions:

1. Assume the time domain is Gaussian
   - Then the frequency domain is Gaussian.
   - Relax the assumption a little to allow coloured noise.
2. Assume the frequency domain is Gaussian.
   - Hence, any given time bin is Gaussian.
   - As a whole the time domain is not Gaussian.

The second of these is closer to reality.

- It deteriorates with wider DFT bins.

Homomorphism revisited

I said that signal and noise are not additive in the power domain.

- In fact, the variances add in the power domain.
- We tend to work in the power domain because of this.

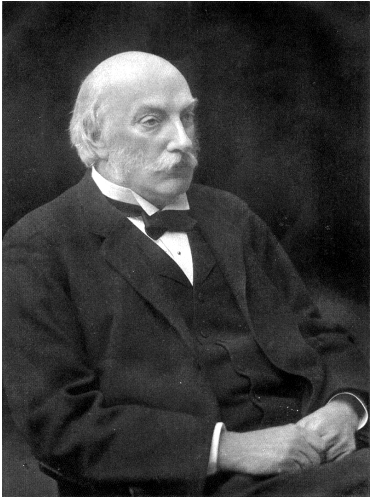
Complex Gaussian

Carl

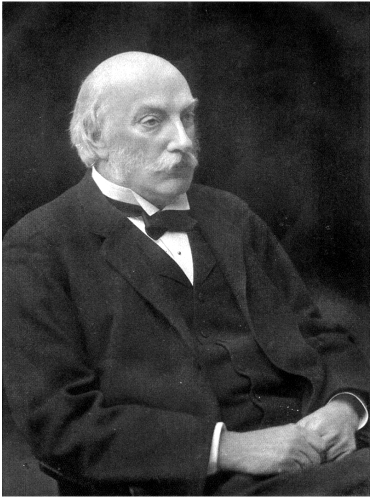
Carl

Carl Friedrich Gauss
1777–1855

John



John



John William Strutt
3rd Baron Rayleigh
1842–1919

Complex Gaussian

Consider a complex number, $\mathfrak{x} = x_{\mathfrak{R}} + i x_{\mathfrak{I}}$, where $i = \sqrt{-1}$. If we consider the components $x_{\mathfrak{R}}$ and $x_{\mathfrak{I}}$ to be zero mean i.i.d.[1] Gaussian RVs, their joint PDF is

$$
\begin{aligned}
p\left(x_{\mathfrak{R}}, x_{\mathfrak{I}} \mid \sigma\right) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{{x_{\mathfrak{R}}}^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{{x_{\mathfrak{I}}}^2}{2\sigma^2}\right) \\
p\left(\mathfrak{x} \mid \sigma\right) &= \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|\mathfrak{x}|^2}{2\sigma^2}\right).
\end{aligned}
$$

[1] Independent and Identically Distributed

Variance

The expectation of the squared magnitude, rather than the square of the variable, is pertinent:

$$\begin{aligned}\mathbb{E}\left(|\mathfrak{x}|^2\right) &= \mathbb{E}\left({x_{\mathfrak{R}}}^2 + {x_{\mathfrak{I}}}^2\right),\\ &= \mathbb{E}\left({x_{\mathfrak{R}}}^2\right) + \mathbb{E}\left({x_{\mathfrak{I}}}^2\right),\\ &= 2\sigma^2.\end{aligned}$$

This motivates parameterisation in terms of a variance $\upsilon = 2\sigma^2$:

$$p\left(\mathfrak{x} \mid \upsilon\right) = \frac{1}{\pi\upsilon} \exp\left(-\frac{|\mathfrak{x}|^2}{\upsilon}\right).$$

Polar form

If we make the substitutions

$$\alpha = |\mathfrak{x}| = \sqrt{{x_{\mathfrak{R}}}^2 + {x_{\mathfrak{I}}}^2} \qquad \theta = \tan^{-1} \frac{x_{\mathfrak{I}}}{x_{\mathfrak{R}}},$$

so that

$$x_{\mathfrak{R}} = \alpha \cos\theta \qquad x_{\mathfrak{I}} = \alpha \sin\theta,$$

the Jacobian determinant is

$$J(\alpha, \theta) = \begin{vmatrix} \dfrac{\partial x_{\mathfrak{R}}}{\partial \alpha} & \dfrac{\partial x_{\mathfrak{I}}}{\partial \alpha} \\ \dfrac{\partial x_{\mathfrak{R}}}{\partial \theta} & \dfrac{\partial x_{\mathfrak{I}}}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & \sin\theta \\ -\alpha\sin\theta & \alpha\cos\theta \end{vmatrix} = \alpha.$$

Distribution of magnitude

This gives

$$p\left(a, \theta \mid \upsilon\right) = \frac{a}{\pi\upsilon} \exp\left(-\frac{a^2}{\upsilon}\right).$$

Integrating over θ gives,

$$\begin{aligned} p\left(a \mid \upsilon\right) &= \int_{-\pi}^{\pi} d\theta \, \frac{a}{\pi\upsilon} \exp\left(-\frac{a^2}{\upsilon}\right) \\ &= \frac{2a}{\upsilon} \exp\left(-\frac{a^2}{\upsilon}\right) \end{aligned}$$

which is the Rayleigh distribution.

Distribution of power (squared magnitude)

If

$$\mathrm{p} = |\mathfrak{x}|^2 = \mathrm{a}^2,$$

so

$$\mathrm{a} = \sqrt{\mathrm{p}}$$

$$\frac{d\mathrm{a}}{d\mathrm{p}} = \frac{1}{2\sqrt{\mathrm{p}}} = \frac{1}{2\mathrm{a}},$$

then

$$\mathrm{p}\left(\mathrm{p} \mid \upsilon\right) = \frac{1}{\upsilon} \exp\left(-\frac{\mathrm{p}}{\upsilon}\right).$$

which is an exponential distribution.

Three forms

The complex Gaussian leads to three common forms depending on whether one is interested in the distribution of the complex number itself, the magnitude or the squared magnitude:

$$\mathfrak{p}\left(\mathfrak{x} \mid \upsilon\right) = \frac{1}{\pi\upsilon} \exp\left(-\frac{|\mathfrak{x}|^2}{\upsilon}\right).$$

$$\mathfrak{p}\left(|\mathfrak{x}| \mid \upsilon\right) = \frac{2|\mathfrak{x}|}{\upsilon} \exp\left(-\frac{|\mathfrak{x}|^2}{\upsilon}\right).$$

$$\mathfrak{p}\left(|\mathfrak{x}|^2 \mid \upsilon\right) = \frac{1}{\upsilon} \exp\left(-\frac{|\mathfrak{x}|^2}{\upsilon}\right).$$

The first is a function of two variables, the latter two are functions of just one variable, and are Rayleigh and exponential distributions respectively.

Distribution of noise variance

Noise observations

Say we can observe N frames of noise, $\{\mathfrak{n}\}_N = \{\mathfrak{n}_1, \mathfrak{n}_2, \ldots, \mathfrak{n}_N\}$, in isolation. We want to estimate the noise variance, ν. This can be written more fully as

$$p\left(\nu_f \mid \{\mathfrak{n}\}_N\right) = \frac{p\left(\nu_f\right) \prod_{i=1}^{N} p\left(\mathfrak{n}_{i,f} \mid \nu_f\right)}{\int_0^\infty d\nu' \, p\left(\nu_f'\right) \prod_{i=1}^{N} p\left(\mathfrak{n}_{i,f} \mid \nu_f'\right)},$$

Hereafter we drop the frame subscripts for simplicity.

Gaussian with weak prior

- The likelihood terms are Gaussian.

$$p\left(n_i \mid \nu\right) = \frac{1}{\pi\nu} \exp\left(-\frac{|n_i|^2}{\nu}\right).$$

- We can choose the priors to be weak (uninformative).

$$p\left(\nu\right) \propto \nu^{-1}.$$

So, the numerator and denominator are of the form

$$\begin{aligned} p\left(\{\mathbf{n}\}_N \mid \nu\right) p\left(\nu\right) &= \nu^{-1} \prod_{i=1}^{N} \frac{1}{\pi\nu} \exp\left(-\frac{|n_i|^2}{\nu}\right) \\ &= \pi^{-N} \nu^{-N-1} \exp\left(-\frac{\sum_{i=1}^{N} |n_i|^2}{\nu}\right) \end{aligned}$$

Inverse Gamma distribution

At this point, it helps to know that the standard form of an inverse gamma distribution is:

$$p\left(x \mid \alpha, \beta\right) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right),$$

which comes from the fact that

$$\int_0^{\infty} dx\, x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right) = \frac{\Gamma(\alpha)}{\beta^{\alpha}}.$$

This gives the answer to the integral that we need.
http://en.wikipedia.org/wiki/Inverse_gamma

Posterior distribution

The posterior distribution of the noise variance is an inverse gamma distribution with parameters:

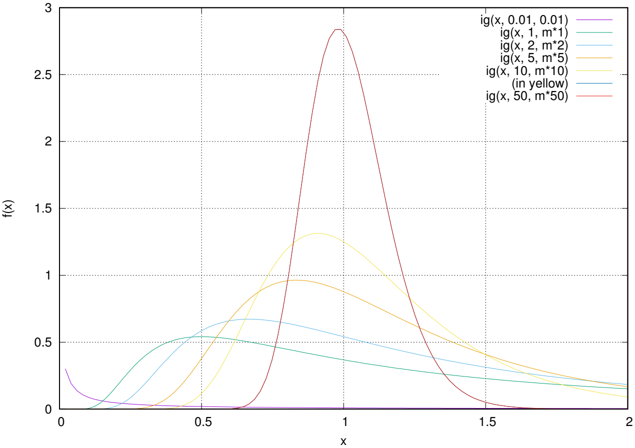
$$\alpha_\nu = N$$
$$\beta_\nu = \sum_{i=1}^{N} |n_i|^2.$$

The MAP estimate, $\hat{\nu}$, of ν is then

$$\hat{\nu} = \frac{\beta_\nu}{\alpha_\nu + 1}.$$

The expectation of ν is

$$\mathbb{E}\left(\nu\right) = \frac{\beta_\nu}{\alpha_\nu - 1}.$$



Inverse gamma distributions for different values of N.