
Everything you wanted to know about Algonquin but were afraid to ask (AKA The tricks conveniently left out of a NIPS paper)

Ron Weiss
Algonquin?

- Model based speech enhancement
- Denoise speech signal corrupted by non-stationary noise
- Needs a prior speech model trained on clean speech
- Can learn noise model directly from noisy signal with EM!
How to log spectra combine?

\[ Y(f) = S(f) + N(f) \]
\[ |Y(f)|^2 = |S(f)|^2 + |N(f)|^2 + 2\text{Re}(N(f)\ast S(f)) \]
\[ = |S(f)|^2 + |N(f)|^2 + E \]

Let \( y = \log |Y(f)|^2 \)

\[ e^y = e^s + e^n + \psi = e^s(1 + e^{n-s}) + \psi \]
\[ y = s + \log(1 + e^{n-s}) + \psi' \]
The Probability Model

Scatter plot of \((x_i - \Delta, n - \Delta)\) where \(\Delta = 8.5 - y_i\), for filter bank 6 at 20dB

Model error due to phase cancellation as Gaussian noise:

\[
p(y|s, n) = \mathcal{N}(y; s + \log(1 + e^{n-s}), \Psi)
\]
The Probability Model

\[ p(y_d = 7|s_d, n_d) \text{ using Algonquin "interaction likelihood"} \]

\[ p(y_d = 7|s_d, n_d) \text{ using max approximation (var = 0.1)} \]
The rest of the probability model

• Separate GMMs for speech and noise
• assume dimensions are independent (i.e. diagonal covariance)
• no temporal dynamics

\[
p(y, s, n, c_s, c_n) = p(y|s, n)p(s|c_s)p(c_s)p(n|c_n)p(c_n)
\]
\[
p(y|s, n) = \mathcal{N}(y; s + \log(1 + e^{n-s}), \Psi)
\]
\[
p(s|c_s)p(c_s) = \mathcal{N}(s; \mu_{c_s}^s, \sigma_{c_s}^s)\pi_{c_s}
\]
\[
p(n|c_n)p(c_n) = \mathcal{N}(n; \mu_{c_n}^n, \sigma_{c_n}^n)\pi_{c_n}
\]
The probability model in pictures

Speech prior $f(x,n) = p(x)$

Noise prior $f(x,n) = p(n)$

Joint distribution $f(x,n) = p(x,n,y_{obs})$
How do we reconstruct the clean signal?

- Our good friend MMSE:

\[
\hat{s} = \int sp(s|y) \, dx = \sum_{c_s} p(c_s) \mu_{c_s}^s \quad \text{(if } p(s|y) \text{ is a mixture of Gaussians)}
\]

- But the posterior \( p(s|y) \), isn’t Gaussian due to non-linear likelihood
Linearization

- Lets linearize the likelihood using 2nd order vector Taylor series!
  - Review: \( f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(a)(x - a)^n \)
  - only want 2 terms: \( f(x) \approx f(a) + f'(a)(x - a) \)
- Let \( g(s, n) = s + \log(1 + e^{n-s}) \), and \( z = [s; n] \)

\[
p(y|n, s) \approx p_l(y|n, s) = \mathcal{N}(y; g(z_0) + g'(z_0)(z - z_0), \Psi)
\]
Lets iterate!

Iteratively update \( z_0 \)
Another Parametrization

• Now we have a Gaussian joint probability, so the posterior is a GMM

• “The form of the joint probability does not allow us to directly read off the mode of the distribution and the marginal”

• “The mode of the posterior $p_l(s, n|y)$ is not coincident with the modes of the priors or the interaction likelihood ($p_l(y|n, s)$)”

$$p_l(s, n, c_s, c_n|y) \approx q(s, n, c_s, c_n) = q(s, n|c_s, c_n)q(c_s, c_n)$$

$$q(s, n|c_s, c_n) = \mathcal{N}([s; n]; [\eta_s; \eta_n], \Phi)$$
Variational approximation

- Find the parameters of $q$ in the standard variational way. Minimize the KL divergence (or equivalently maximize $\log p(y) - KL$ divergence) between $p_l$ and $q$:

$$KL(p||q) = \sum_{c_s} \sum_{c_n} \int_s \int_n q(s,n,c_s,c_n) \log \frac{q(s,n,c_s,c_n)}{p(s,n,c_s,c_n|y)}$$

$$\log p(y) - KL(p||q) = C - \sum_{c_s} \sum_{c_n} \int_s \int_n q(s,n,c_s,c_n) \log \frac{p(s,n,c_s,c_n,y)}{q(s,n,c_s,c_n)}$$

- Take derivatives with respect to each parameter of $q$, set to zero and solve to find new parameters in terms of the old ones.

- Can now find MMSE estimate for $s$: $\hat{s} = \sum_{c_s} p(c_s) \eta_{c_s}^s$
So...

- Algonquin is the greatest thing since sliced bread
- Better approximation to the conditional likelihood of $y$ given $s, n$ than the max approximation
- It's also a good bit slower (at least when you code it in MATLAB)
  - The new parametrization couples the means of the GMMs for $s$ and $n$
  - $\eta$ needs to be recalculated at every iteration
- Is it really worth it?
Max separation

model 1

model 2

mixture

speaker 1 reconstruction (max)

speaker 2 reconstruction (max)

Algonquin separation