Signal Enhancement

Bastiaan Kleijn
KTH School of Electrical Engineering
Stockholm

Overview

- Motivation / Introduction
- Basic noise- and signal (speech) power spectrum estimation
- Estimating noise-free speech given noise/speech models:
  - Linear estimators: Wiener filter in various flavors
  - General estimators: Wiener filter and other estimators
- Probabilistic estimation of noise and speech models
- Performance

Problem Definition

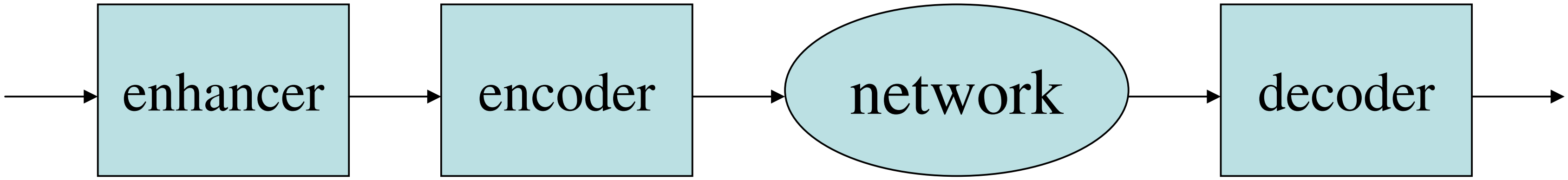
- Signal corrupted by additive noise
  - $Y^k = X^k + W^k$
  - X^k and W^k statistically independent
- Estimate noise-free signal X^k

Motivation for Enhancement

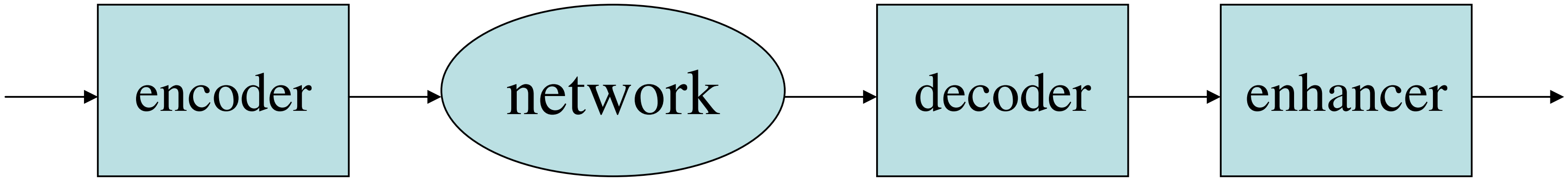
- Plain old telephone service (POTS)
  - Generally low acoustic background noise level
    - Phone booth
    - Home
- Modern networks
  - Often high acoustic background noise level
    - Mobile phones
    - Computer as phone
- Complication:
  - Not difficult to improve SNR
  - Difficult to obtain enhanced signal that *sounds* more pleasant than noisy signal

Location in Network

- Input based
  - Obvious location
  - Best performance, in commercial use



- Output based
  - Quality resides with purchaser of device

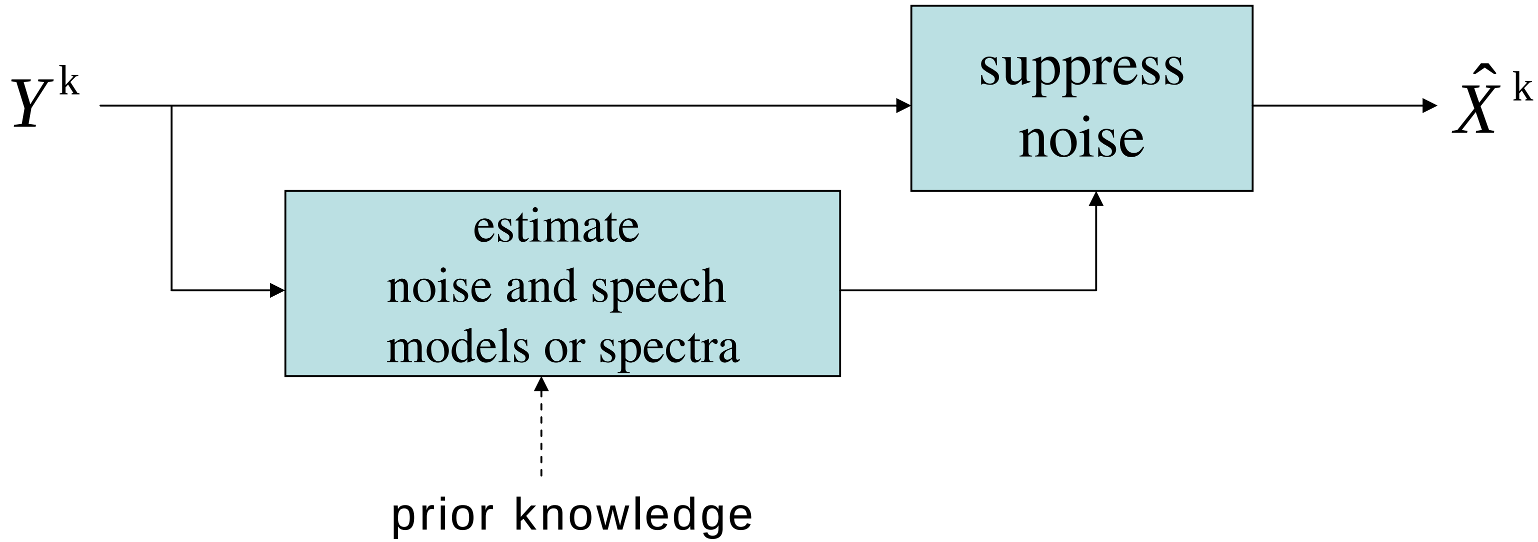


Single- and Multi-Channel Enhancement

- Single-channel
  - Beneficial if no control of input device
  - Inexpensive
  - Now common
- Multi-channel
  - Adaptive noise cancellation
    - Assumes noise reference available
  - Adaptive beam forming
    - Physical model of environment
  - Blind source separation
    - No physical model
    - Assumes convolutive mixing, independent sources
    - Outputs are filtered versions of original

Single-Channel Architecture

- General algorithmic steps
  1. Estimate noise *and speech* model (variance, power spectrum / AR parameters)
     - May exploit prior knowledge of signal structure
  2. Estimate clean speech signal $\hat{X}^k$
     - Exploit speech or speech/noise model



Overview

- Motivation / Introduction
- Basic noise- and speech power spectrum estimation
- Estimating noise-free speech given noise/speech models:
  - Linear estimation: Wiener filter in various flavors
  - General estimation: Wiener filter and other estimators
- Probabilistic estimation of noise and speech models
- Performance

Voice-Activity Based Noise-PS Estimation

- Objective: to estimate power spectrum of noise P_W^k
- Algorithm for each block:
  1. Voice activity detection
     Based on spectral slope, signal power, autocorrelation, etc.
  2. If no speech present compute periodogram
  3. Average periodograms over suitable interval
  4. Result is power spectral estimate of noise $\hat{P}_W^k$
- Main weaknesses:
  - Voice activity detection notoriously unreliable
    - Operates on noisy signal
  - Assumes noise is stationary

Noise PS Estimation by Minimum Statistics

R. Martin, Aachen -> Bochum

- Objective: to estimate power spectrum of noise P_W^k

- Algorithm for each block:
  1. Compute periodogram $|(Fy^k)(\mathrm{m})|^2$ of noisy signal y^k
  2. Smooth periodogram across time (and frequency)
  3. Add new periodogram to stored set of last L periodograms
  4. Remove oldest periodogram from set of last L periodograms
  5. For each freq bin, find min value in periodogram set
  6. Compensate for estimation bias, get $\hat{P}_W^k$

- Main weakness:
  - High computational effort
  - Requires near-stationarity of noise
    - Increase set cardinality -> more reliable but slower adaptation

Noise PS Estimation by Minimum Statistics

- Frame = 15 ms, 100 frames=1.5 s, k=25~ 800 Hz
- **Noise power spectral density estimation based on optimal smoothing and minimum statistics**
  Martin, R.; Speech and Audio Processing, IEEE Transactions on,Volume 9, Issue 5, July 2001 Page(s):504 - 512

Quantile Based Noise PS Estimation

- Objective: to estimate power spectrum of noise P_W^k

- Algorithm for each block:
  1. Compute periodogram $|(Fy^k)(\mathrm{m})|^2$ of noisy signal y^k
  2. Add new periodogram to stored set of last L pgrams
  3. Remove oldest periodogram from set of last L pgrams
  4. For each freq bin, find q^{th} quantile in periodogram set: $\hat{P}_W^k$

- Main weakness:
  - High computational effort
  - Requires near-stationarity of noise
    - increase set cardinality -> more reliable but slower adaptation

Quantile Based Noise PS Estimation

- **Quantile based noise estimation for spectral subtraction and Wiener filtering**
  Stahl, V.; Fischer, A.; Bippus, R.;Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on,Volume 3, 5-9 June 2000 Page(s):1875 - 1878 vol.3

Speech PS Estimation by Subtraction

- Objective: to estimate speech power spectrum P_X^k
- Solution: ad-hoc but simple

- Spectral subtraction (ad hoc):

$$\sqrt{\hat{P}_X^k(m)} = \max(0, |(Fy^k)(m)| - \sqrt{\hat{P}_W^k(m)})$$

- Power spectral subtraction (logical if signals uncorrelated):

$$\hat{P}_X^k(m) = \max(0, |(Fy^k)(m)|^2 - \hat{P}_W^k(m))$$

- Low signal magnitude: estimates poor ⇨ musical noise

Overview

- Motivation / Introduction
- Basic noise- and speech power spectrum estimation
- Estimating noise-free speech given noise/speech models:
  - Linear estimation: Wiener filter in various flavors
  - General estimation: Wiener filter and other estimators
- Probabilistic estimation of noise and speech models
- Performance

Speech Estimation Problem

- Given
  - Speech power spectrum $\hat{P}_X^k$
  - Noise power spectrum $\hat{P}_W^k$
  - Noisy speech y^k

- We want MMSE estimate

$$\underset{v}{\operatorname{argmin}}\, \mathrm{E}[|X^k - v|^2] = \mathrm{E}[X^k]$$

Ad-hoc Noise Suppression

- We have
  - Estimate speech power spectrum $\hat{P}_X^k$
  - Estimate noise power spectrum $\hat{P}_W^k$
  - Noisy speech y^k
- We want $\mathrm{E}[X^k]$

- *Ad-hoc* solution: ignore phase, Spectral Subtraction:

$$(\mathrm{F}\hat{\mathrm{x}}^{\mathrm{k}})(\mathrm{m}) = \sqrt{\frac{\hat{P}_X^k(m)}{\hat{P}_Y^k(m)}}\ (\mathrm{Fy}^{\mathrm{k}})(\mathrm{m})$$

$$\hat{x}^k = F^H(\mathrm{diag}\,\hat{P}_X^k)(\mathrm{diag}\,\hat{P}_Y^k)^{-1}\ F\ y^k$$

- Not based on MMSE criterion but ad-hoc solution is essentially what we get from Wiener filter

Formal Solution: Wiener Filter

- *Operator* that estimates speech from noisy speech

- Follows from either of two equivalent assumptions:
  1. MMSE *linear* estimate
  2. Best estimate if speech and noise have *Gaussian distribution*

MMSE Linear Estimator H

- Linear estimate $\hat{X}^k = H\, Y^k$
- MSE that retains some noise:

$$\eta = \mathrm{E}\left\| (X^k + \varepsilon W^k) - \hat{X}^k \right\|_2$$

$$= \mathrm{E}\left[\mathrm{tr}\left[\left((X^k + \varepsilon W^k) - \hat{X}^k \right) \left((X^k + \varepsilon W^k) - \hat{X}^k \right)^H \right] \right]$$

- Find optimal linear estimator H
  - Assume covariance matrices of R_W and R_X known

Wiener Filter that Minimizes MSE

why we used the "trace" notation

- MSE criterion:

$$\eta = \mathrm{E}\left[\mathrm{tr}\left[(X^k + \varepsilon W^k - HY^k)\,(X^k + \varepsilon W^k - HY^k)^H\right]\right]$$

$$= \mathrm{E}\left[\mathrm{tr}\left[(H - I)X^k + (H - \varepsilon I)W^k)\,((H - I)X^k + (H - \varepsilon I)W^k\right]^H\right]$$

$$= \mathrm{tr}\left[(H - I)R_X(H - I)^H\right] + \mathrm{tr}\left[(H - \varepsilon I)R_W(H - \varepsilon I)^H\right]$$

- Differentiate to H_{ij} ; set to zero; solve;

Wiener filter: $H = (R_X + \varepsilon R_W)\,(R_X + R_W)^{-1}$

Wiener Filter that Minimizes MS *Distortion*

Y. Ephraim

- Estimation error is $$X^k - \hat{X}^k = X^k - HY^k$$ $$= (H-I)X^k - HW^k$$
  - Distortion: $(H-I)X^k$
  - Residual noise: HW^k

- Alternative: minimize distortion given residual noise (or vice versa)

$$\eta = \mathrm{E}\left\|(H-I)X^k\right\|_2 + \mu\, \mathrm{E}\, \| HW^k \|_2$$

$$= \mathrm{tr}\left[(H-I)R_X(H-I)^H\right] + \mu\, \mathrm{tr}\left[(HR_W H^h\right]$$

- Solution: $H = R_X\ (R_X + \mu\ R_W)^{-1}$
  - Each μ corresponds to particular residual noise level

Wiener Filter: Cyclic Approximation

- Discrete Fourier transform is a matrix F
  - inverse Fourier transform is F^H; that is $F^H F = I$

- Properties of R_X and R_W
  - stationary signals: Toeplitz and symmetric
  - periodic stationary signals: circulant and symmetric
  - Fourier T diagonalizes circulant symmetric matrices: Diagonal of matrix is diag(*power spectral density)*
- Under periodic approximation:

$$
\begin{aligned}
FHF^H &= F(R_X + \varepsilon R_W)F\ F^H(R_X + R_W)^{-1}F^H \\
&= (FR_X F^H + \varepsilon FR_W F^H)\ (FR_X F^H + FR_W F^H)^{-1} \\
&\approx (\operatorname{diag} P_X + \varepsilon \operatorname{diag} P_W)(\operatorname{diag} P_X + \operatorname{diag} P_W)^{-1}
\end{aligned}
$$

Wiener Filter and Subspace Methods

Y. Ephraim

- Assume white noise (pre- and post- filter if it is not)
  - Then R_W scaled identity matrix
- Note

$$R_Y = R_X + R_W = R_X + \Lambda_W = U^H \Lambda_X U + U^H R_W U$$

- Retain only subspace (spanned by rows of U) corresponding to large eigenvalues of Λ_X
  - Somewhat ad-hoc

remains diagonal for white noise

Wiener Filter and Kalman Filter

- Kalman filter is a time-varying filter
  - Model of signal and noise known; state-space formulation:

$$x^k(n+1) = Ax^k(n) + Gv^k(n)$$

$$y(n) = Bx^k(n) + w^k(n)$$

  - Objective: find MMSE estimate of $x^k(n)$ given $\cdots, y(n-2), y(n-1), y(n)$ and state-space model
  - Main difference to Wiener filter: causality!
    - Causality reduces performance
    - Can handle time-variant speech/noise model
    - Low number of parameters: perceptual weighing helps

- Kalman smoother: allows delay ⇨
  - Converges to Wiener filter performance
  - Speech: small delay gives near-optimal performance

On causal algorithms for speech enhancement Grancharov, V.; Samuelsson, J.; Kleijn, B.;Audio, Speech and Language Processing, IEEE Transactions on [see also Speech and Audio Processing, IEEE Transactions on] Volume 14, Issue 3, May 2006 Page(s):764 - 773

Overview

- Motivation / Introduction
- Basic noise- and speech power spectrum estimation
- Estimating noise-free speech given noise/speech models:
  - Linear estimation: Wiener filter in various flavors
  - General estimation: ML estimator, Wiener filter, etc.
  - Estimation based on distribution of models
- Probabilistic estimation of noise and speech models
- Performance

Side-Step: ML Estimate

- Max likelihood estimate:

$$\hat{x}^k = \arg\max_{x^k} p_{Y|X}(y^k \mid x^k)$$
$$= \arg\max_{x^k} p_{W|X}(y^k - x^k \mid x^k)$$
$$= \arg\max_{x^k} p_W(y^k - x^k)$$
$$= y^k$$

- Conclusion: *ML speech estimate does not reduce noise*!

MMSE Estimator

- MMSE estimation:

$$\hat{x}^k = \arg\min_{v^k} \mathrm{E}\left[(X^k - v^k)^2 \mid Y^k = y^k \right]$$

$$= \mathrm{E}\left[X^k \mid y^k \right]$$

$$= \int x^k p_{X|Y}(x^k \mid y^k) dx^k$$

$$= \int x^k p_{Y|X}(y^k \mid x^k) p_X(x^k) / p_Y(y^k) dx^k$$

Gaussian Assumption for PDFs

- Speech and noise have Gaussian distribution:

$$p_X(x^k) = \frac{1}{\sqrt{(2\pi)^k \det(R_X)}} \exp(-x^{kT} R_X^{-1} x^k / 2)$$

$$p_W(w^k) = \frac{1}{\sqrt{(2\pi)^k \det(R_W)}} \exp(-w^{kT} R_W^{-1} w^k / 2)$$

$$p_Y(y^k) = \frac{1}{\sqrt{(2\pi)^k \det(R_X + R_W)}} \exp(-y^{kT} (R_X + R_W)^{-1} y^k / 2)$$

- Then:

$$p_W(y^k - x^k) = \frac{1}{\sqrt{(2\pi)^k \det(R_Y - R_X)}} \exp(-(y^k - x^k)^T (R_Y - R_X)^{-1} (y^k - x^k) / 2)$$

MMSE Estimator

- MMSE estimation:

$$\hat{x}^k = \arg\min_{v^k} \mathrm{E}\left[(X^k - v^k)^2 \mid Y^k = y^k \right]$$

$$= \mathrm{E}\left[X^k \mid y^k \right]$$

$$= \int x^k p_{X|Y}(x^k \mid y^k) dx^k$$

$$= \int x^k p_{Y|X}(y^k \mid x^k) p_X(x^k) / p_Y(y^k) dx^k$$

- Next, we work out the argument of the integral

MMSE Estimator: the Gaussian I

- Rewrite density (y^k is constant; complete the square):

$$p_W(y^k - x^k)\, p_X(x^k) / p_Y(y^k) =$$

$$= C \exp\left(\frac{-(y^k - x^k)^T R_W^{-1}(y^k - x^k) + x^{kT} R_X^{-1} x^k}{2}\right)$$

$$= C' \exp\left(\frac{-x^{kT}(R_X^{-1} + R_W^{-1})x^k + 2x^{kT} R_W^{-1} y^k}{2}\right)$$

$$= C' \exp\left(\frac{-x^{kT}(R_X^{-1} + R_W^{-1})x^k + 2x^{kT} R_W^{-1} y^k}{2}\right)$$

$$= C' \exp\left(\frac{-x^{kT}(R_X^{-1} + R_W^{-1})x^k + 2x^{kT}(R_X^{-1} + R_W^{-1})(R_X^{-1} + R_W^{-1})^{-1} R_W^{-1} y^k}{2}\right)$$

$$= C' \exp\left(\frac{-(z^k - x^k)^T (R_X^{-1} + R_W^{-1})(z^k - x^k)}{2}\right)$$

$$= p_U(z^k - x^k)$$

where:

$$z^k = (R_X^{-1} + R_W^{-1})^{-1} R_W^{-1}\, y^k = R_X (R_W + R_X)^{-1} y^k = R_X R_Y^{-1} y^k$$

MMSE Estimator: the Gaussian II

- Back to the MMSE estimate:

$$\begin{aligned}
\hat{x}^k &= \mathrm{E}\left[X^k \mid y^k \right] \\
&= \int x^k p_U(z^k - x^k) dx^k \\
&= \int (x^k - z^k) p_U(z^k - x^k) dx^k + z^k \int p_U(z^k - x^k) dx^k \\
&= z^k \\
&= R_X R_Y^{-1} y^k
\end{aligned}$$

- Is linear!
- Is the Wiener filter!

MMSE Estimator

- MMSE estimation:

$$\hat{x}^k = \arg\min_{v^k} \mathrm{E}\left[(X^k - v^k)^2 \mid Y^k = y^k \right]$$

$$= \mathrm{E}\left[X^k \mid y^k \right]$$

$$= \int x^k p_{X|Y}(x^k \mid y^k) dx^k$$

$$= \int x^k p_{Y|X}(y^k \mid x^k) p_X(x^k) / p_Y(y^k) dx^k$$

General MMSE Estimator I

- Variants on criterion:
  - Gaussian but MMSE on amplitude only
    - **Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator,** Ephraim, Y.; Malah, D.;Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on, Volume 32, Issue 6, Dec 1984, Page(s):1109 - 1121

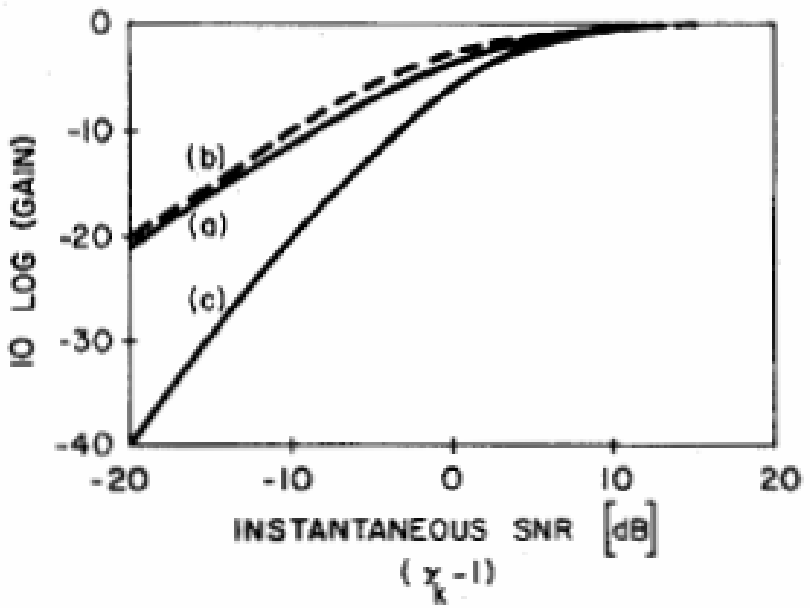


Fig. 6. Gain curves describing (a) MMSE gain function $G_{MMSE}(\xi_k, \gamma_k)$ defined by (7) and (14), with $\xi_k = \gamma_k - 1$, (b) "spectral subtraction" gain function (46) with $\beta = 1$, and (c) Wiener gain function $G_W(\xi_k, \gamma_k)$ (15) with $\xi_k = \gamma_k - 1$.

Overview MMSE Estimator

- Variants on speech distribution:
  - Gaussian but MMSE on amplitude only
    - Similar effect as estimating $X^k + \varepsilon W^k$
  - Super-Gaussian models
    - Gamma distribution of DFT coefficients
    - Minor improvement

Overview

- Motivation / Introduction
- Basic noise- and speech power spectrum estimation
- Estimating noise-free speech given noise/speech models:
  - Linear estimation: Wiener filter in various flavors
  - General estimation: ML estimator, Wiener filter, etc.
  - Estimation based on distribution of models
- Probabilistic estimation of noise and speech models
- Performance

Distribution of Models

- Model distributions
  - Speech/noise models each have probability given observed data $p_{\Theta|Y^k}(\theta \mid y^k)$
  - The MMSE is averaged over models:

$$\hat{x}^k = \arg\min_{v^k} \mathrm{E}\left[\| X^k - v^k \|^2 \mid y^k\right] = \mathrm{E}\left[X^k \mid y^k\right]$$

$$= \int x^k p_{X^k|Y^k}(x^k \mid y^k) dx^k$$

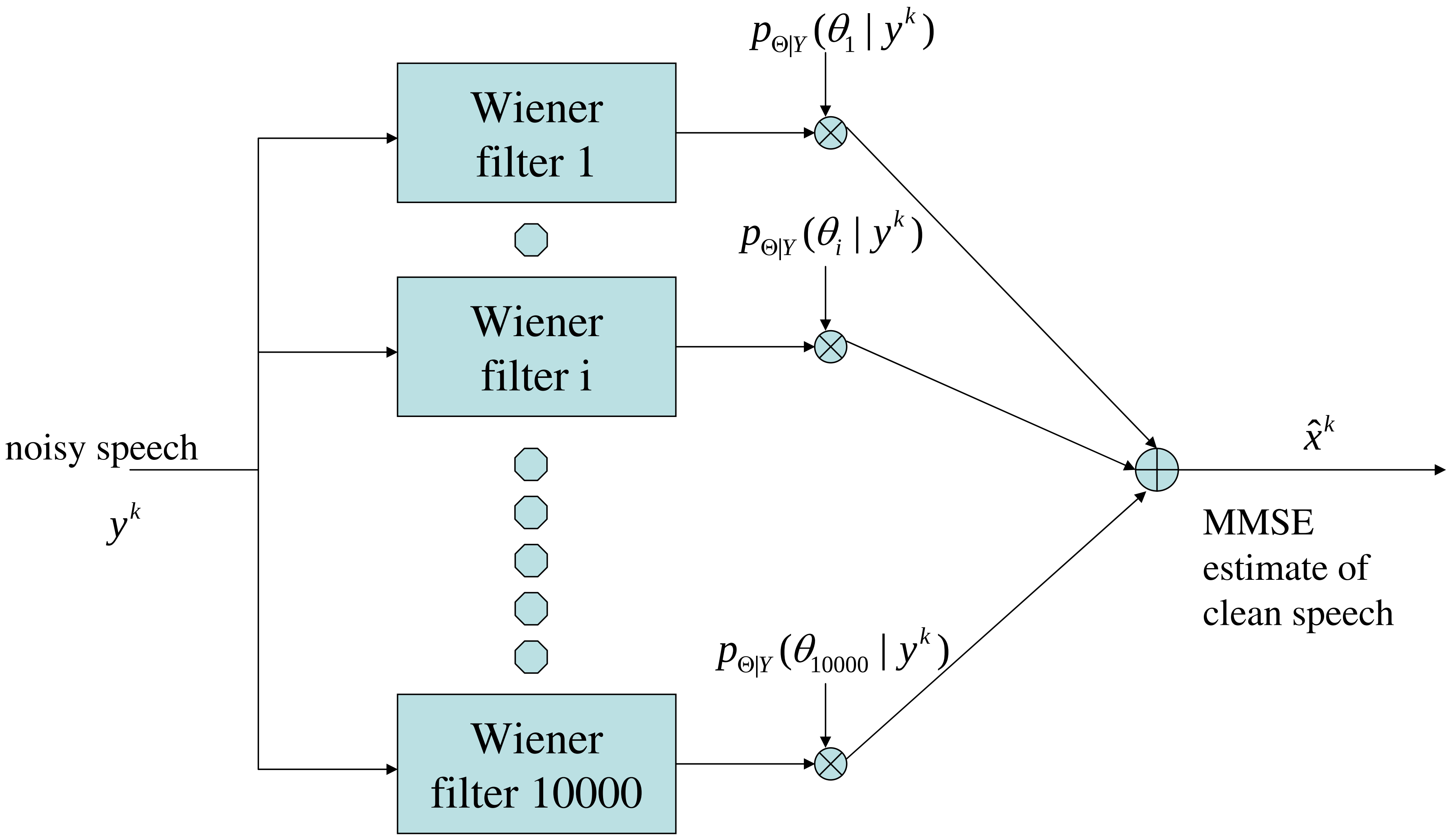
$$= \int x^k \int p_{X^k|Y^k,\Theta}(x^k \mid y^k, \theta)\; p_{\Theta|Y^k}(\theta \mid y^k) d\theta \; dx^k$$

$$= \int \int x^k p_{X^k|Y^k,\Theta}(x^k \mid y^k, \theta) dx^k \quad p_{\Theta|Y^k}(\theta \mid y^k) d\theta$$

$$= \int R_{X^k}(\theta) R_{Y^k}^{-1}(\theta) y^k \; p_{\Theta|Y^k}(\theta \mid y^k) d\theta$$

Gaussian

Codebook of Models I



Codebook of Models II

- Consider a codebook with 1000 speech spectra and 10 noise spectra; Gaussian speech & noise assumption
- 10.000 speech+noise covariance matrices / spectra
- Each combination has a corresponding Wiener filter
- Each combination has a probability given the data
- Compute speech estimate as weighted sum of Wiener filters operating on noisy input

prevents ambiguity

The Details for the Distribution Case

- Gaussian speech assumption:

$$\begin{aligned}
\hat{x}^k &= \mathrm{E}[X^k \mid y^k] \\
&= \int x^k p_{X|Y}(x^k \mid y^k) dx^k \\
&= \int x^k \left(\int p_{X|Y,\Theta}(x^k \mid y^k, \theta) p_\Theta(\theta^k \mid y^k) d\theta \right) dx^k \\
&= \int \left(\int x^k p_{X|Y,\Theta}(x^k \mid y^k, \theta) dx^k \right) p_{\Theta|Y}(\theta \mid y^k) d\theta^k \\
&= \int \left(\int x^k p_{X|Y,\Theta}(x^k \mid y^k, \theta) dx^k \right) \frac{p_{Y|\Theta}(y^k \mid \theta) p_\Theta(\theta)}{\int p_{Y|\Theta}(y^k \mid \theta) p_\Theta(\theta) d\theta} d\theta \\
&= \int \left(R_X(\theta) R_Y^{-1}(\theta) y^k \right) \frac{p_{Y|\Theta}(y^k \mid \theta) p_\Theta(\theta)}{\int p_{Y|\Theta}(y^k \mid \theta) p_\Theta(\theta) d\theta} d\theta \\
&= \int R_X(\theta) R_Y^{-1}(\theta) \frac{p_{Y|\Theta}(y^k \mid \theta) p_\Theta(\theta)}{\int p_{Y|\Theta}(y^k \mid \theta) p_\Theta(\theta) d\theta} d\theta \quad y^k
\end{aligned}$$

from Gaussian assumption

must be measured or postulated

MMSE for Distribution of Models

- Model distributions:

$$
\begin{aligned}
\hat{x}^k &= \mathrm{E}\left[X^k \mid y^k \right] \\
&= \int x^k p_{X|Y}(x^k \mid y^k) dx^k \\
&= \int x^k \left(\int p_{X|Y,\Theta}(x^k \mid y^k, \theta) p_{\Theta}(\theta^k \mid y^k) d\theta \right) dx^k \\
&= \int \left(\int x^k p_{X|Y,\Theta}(x^k \mid y^k, \theta) dx^k \right) p_{\Theta|Y}(\theta \mid y^k) d\theta
\end{aligned}
$$

- We simply average the estimates

- If each $p_{X|Y,\Theta}(x^k \mid y^k, \theta)$ corresponds to Gaussian noise and speech models, then we average the corresponding Wiener filters!

MMSE for Distribution of Models

- Still missing:
  - The density $p_{\Theta|Y}(\theta_1 \mid y^k)$

Overview

- Motivation / Introduction
- Basic noise- and speech power spectrum estimation
- Estimating noise-free speech given noise/speech models:
  - Linear estimation: Wiener filter in various flavors
  - General estimation: Wiener filter and other estimators
- Probabilistic estimation of noise and speech models
- Performance

Prob Noise and Speech Model Estimation

- Approach I: for single model case
  - Find one ‘optimal’ speech and one ‘optimal’ noise model
    - Spectral subtraction
    - ML estimate of
    - MAP estimate of
    - MMSE estimate of

$$\theta = \{\theta_{\text{speech}}, \theta_{\text{noise}}\}$$

  - Find MMSE estimate of speech given this combination
  - *(Remember ML estimate speech not sensible*)
  - Advantage: estimate based on true speech and noise models
  - Disadvantage: larger MSE

- Approach II: for distribution of models
  - Find posterior distribution of models $p_{\Theta}(\theta \mid y^k)$
  - Find MMSE estimate of speech $\mathrm{E}[X^k \mid y^k]$
    - use posterior distribution
  - Disadvantage: output does not have to be “true speech”
  - Advantage: smaller MSE

MAP and ML Estimation of θ

must be measured or postulated

- Maximum *posterior* probability (MAP):

$$\arg\max_{\theta} p_{\Theta|Y}(\theta \mid y^k) = \arg\max_{\theta} \frac{p_{Y|\Theta}(y^k \mid \theta)\, p_{\Theta}(\theta)}{p_Y(y^k)}$$

$$= \arg\max_{\theta} p_{Y|\Theta}(y^k \mid \theta)\, p_{\Theta}(\theta)$$

- ML: *prior* probability constant (= θ is deterministic)

$$\arg\max_{\theta} p_{\Theta|Y}(\theta \mid y^k) = \arg\max_{\theta} \frac{p_{Y|\Theta}(y^k \mid \theta)\, p_{\Theta}(\theta)}{p_Y(y^k)}$$

$$= \arg\max_{\theta} p_{Y|\Theta}(y^k \mid \theta)$$

Gaussian Assumption

- Speech and noise satisfy AR model:

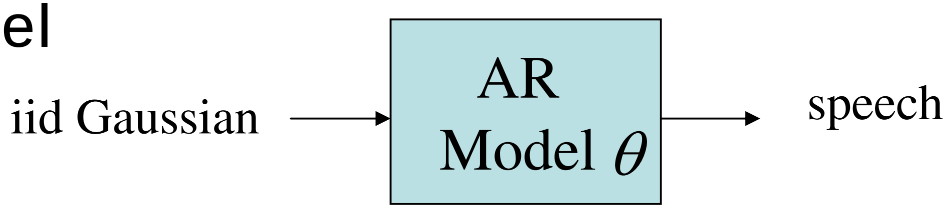
$$p_{X|\Theta}(x^k \mid \theta_{\text{speech}}) = \frac{1}{\sqrt{(2\pi)^k \det(R_X)}} \exp(-x^{kT} R_X^{-1}(\theta_{\text{speech}}) x^k / 2)$$

$$p_{W|\Theta}(w^k \mid \theta_{\text{noise}}) = \frac{1}{\sqrt{(2\pi)^k \det(R_W)}} \exp(-w^{kT} R_W^{-1}(\theta_{\text{noise}}) w^k / 2)$$

$$p_{Y|\Theta}(y^k \mid \theta) = \frac{1}{\sqrt{(2\pi)^k \det(R_X + R_W)}} \exp(-y^{kT} (R_X + R_W)^{-1} y^k / 2)$$

Model Family/Structure: Gaussian Assumption and AR Model

- AR model



- Implication:

$$
\begin{aligned}
R_X &= \mathrm{E}\left[x^k x^{kH}\right] \\
&= \mathrm{E}\left[A e^k e^{kH} A^H\right] \\
&= \sigma^2 A A^H \\
&= R_X(\theta)
\end{aligned}
$$

$A = A(\theta)$

- *A* is Toeplitz, since AR model is a linear filter

- Circulant approximation: $F R_X F^H = \sigma_e^2 \,\mathrm{diag}(P_A)$

MAP and ML Estimation of θ

must be measured or postulated

- Maximum *posterior* probability (MAP):

$$\arg\max_{\theta} p_{\Theta|Y}(\theta \mid y^k) = \arg\max_{\theta} \frac{p_{Y|\Theta}(y^k \mid \theta)\, p_{\Theta}(\theta)}{p_Y(y^k)}$$

$$= \arg\max_{\theta} p_{Y|\Theta}(y^k \mid \theta)\, p_{\Theta}(\theta)$$

- ML: *prior* probability constant (= θ is deterministic)

$$\arg\max_{\theta} p_{\Theta|Y}(\theta \mid y^k) = \arg\max_{\theta} \frac{p_{Y|\Theta}(y^k \mid \theta)\, p_{\Theta}(\theta)}{p_Y(y^k)}$$

$$= \arg\max_{\theta} p_{Y|\Theta}(y^k \mid \theta)$$

Example: Codebook ML and MAP

- ML algorithm
  - For all model combinations $\theta = \{\theta_{\text{speech}}, \theta_{\text{noise}}\}$ evaluate likelihood for y^k
  - Select model with maximum likelihood

- MAP algorithm
  - Presume a prior $p_\Theta(\theta)$
  - For all model combinations $\theta = \{\theta_{\text{speech}}, \theta_{\text{noise}}\}$ evaluate a-posteriori probability for y^k
  - Select model with maximum a-posteriori probability

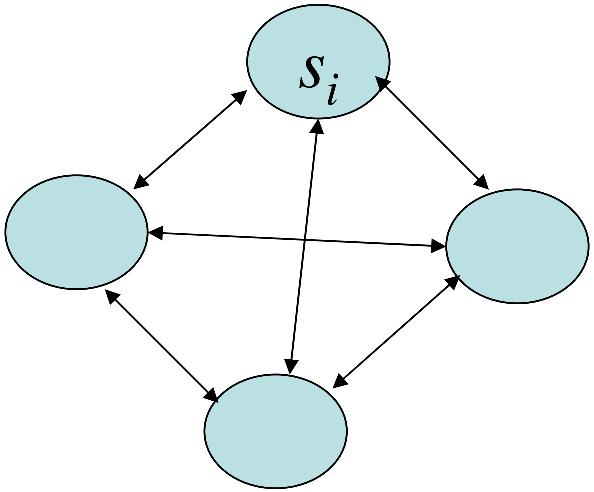
Side-Step: Introducing Memory I

- Not sensible for ML
- Easy for MAP
- Difficult for MMSE

Side-Step: Introducing Memory II

- First-order Markov model

$$p_{S|S,\cdots}(s_i \mid s_{i-1}, s_{i-2}\cdots) = p_{S|S}(s_i \mid s_{i-1})$$



Side-Step: Introducing Memory III

- First-order Markov model

$$\arg\max_{\theta} p_{\Theta|Y}(\theta \mid y_i^k)$$

$$= \arg\max_{\theta} p_{Y|\Theta}(y_i^k \mid \theta)\, p_{\Theta}(\theta) / p(y_i^k)$$

$$= \arg\max_{\theta} p_{Y|\Theta}(y_i^k \mid \theta)\, p_{\Theta}(\theta)$$

$$= \arg\max_{\theta} \sum_{s_i} p_{Y|\Theta,S}(y_i^k \mid \theta, s_i)\, p_{\Theta,S}(\theta \mid s_i) p_S(s_i)$$

$$= \arg\max_{\theta} \sum_{s_{i-1}} \sum_{s_i} p_{Y|\Theta,S}(y_i^k \mid \theta, s_i)\, p_{\Theta,S}(\theta \mid s_i) p_S(s_i \mid s_{i-1}) p_S(s_{i-1})$$

- Use Viterbi algorithm to find optimal sequence

MMSE Estimation of θ

- MMSE estimation of $\theta = \{\theta_{\text{speech}}, \theta_{\text{noise}}\}$
  - Continuous case:

$$\hat{\theta} = \mathrm{E}[\Theta \mid y^k]$$
$$= \int \theta \, p_{\theta|Y}(\theta \mid y^k) d\theta$$
$$= \int \frac{\theta \, p_{Y|\theta}(y^k \mid \theta) p_{\Theta}(\theta)}{p_Y(y^k)} d\theta$$
$$= \frac{\int \theta \, p_{Y|\theta}(y^k \mid \theta) p_{\Theta}(\theta) d\theta}{\int p_{Y|\theta}(y^k \mid \theta) p_{\Theta}(\theta) d\theta}$$

  - MSE must be *reasonable* -> LSF

Codebook MMSE

- MMSE estimation of $\theta = \{\theta_{\text{speech}}, \theta_{\text{noise}}\}$
  - Discrete case: $\hat{\theta} = \mathrm{E}[\Theta \mid y^k]$

$$= \frac{\sum \theta \, p_{Y|\theta}(y^k \mid \theta_i) \, p_{\Theta}(\theta_i)}{\sum p_{Y|\theta}(y^k \mid \theta_i) \, p_{\Theta}(\theta_i)}$$

Mixture Model for Model Parameters

- Mixture prior model: $p_{\Theta}(\theta) = \sum_i c_i p_{\theta,i}(\theta)$

- Simple to combine with:

$$\begin{aligned}
\hat{\theta} &= \mathrm{E}[\Theta \mid y^k] \\
&= \int \theta p_{\theta|Y}(\theta \mid y^k) d\theta \\
&= \int \frac{\theta p_{Y|\theta}(y^k \mid \theta) p_{\Theta}(\theta)}{p_Y(y^k)} d\theta \\
&= \frac{\int \theta p_{Y|\theta}(y^k \mid \theta) p_{\Theta}(\theta) d\theta}{\int p_{Y|\theta}(y^k \mid \theta) p_{\Theta}(\theta) d\theta}
\end{aligned}$$

- May need numerical approximations

Noise and Speech Model Estimation

- Approach I: for single-model case
  - Find one ‘optimal’ speech and one ‘optimal’ noise model
    - Spectral subtraction
    - ML estimate of
    - MAP estimate of $\theta = \{\theta_{\text{speech}}, \theta_{\text{noise}}\}$
    - MMSE estimate of
  - Find MMSE estimate of speech given this combination
  - *(Remember ML estimate speech not sensible*)
  - Advantage: estimate based on true speech and noise models
  - Disadvantage: larger MSE

- Approach II: for distribution of models
  - Find posterior distribution of models $p_{\Theta}(\theta \mid y^k)$
  - Find MMSE estimate of speech $\mathrm{E}[X^k \mid y^k]$
    - use posterior distribution
  - Disadvantage: output does not have to be “true speech”
  - Advantage: smaller MSE

Posterior Distribution

- Posterior distribution in terms of known distributions

must be measured
or postulated

$$p_{\Theta|Y^k}(\theta \mid y^k) = \frac{p_{Y^k|\Theta}(y^k \mid \theta)\, p_\Theta(\theta)}{p_{Y^k}(y^k)}$$

$$= \frac{p_{Y^k|\Theta}(y^k \mid \theta)\, p_\Theta(\theta)}{\int p_{Y^k|\Theta}(y^k \mid \theta)\, p_\Theta(\theta) d\theta}$$

Additional Issues

- Gain:
  - Noise and speech gain varies strongly:
    - Separate scaling for model
- How to obtain models for Θ
  - Codebook
    - Random sampling data base
    - Lloyd algorithm
  - Gaussian mixtures / HMM
    - Expectation maximization (EM) algorithm

Overview

- Motivation / Introduction
- Basic noise- and speech power spectrum estimation
- Estimating noise-free speech given noise/speech models:
  - Linear estimation: Wiener filter in various flavors
  - General estimation: Wiener filter and other estimators
- Probabilistic estimation of noise and speech models
- Performance

Typical Performance

- Typical problem:
  - “Musical” noise
  - Performance became acceptable in commercial applications: 1990-1995

- Performance better for stationary signals

Conclusions

- Motivated by ubiquitous network
- Nice application for estimation theory
- Methods
  - Approach I
    - Find one ‘optimal’ speech and one ‘optimal’ noise model
    - Find MMSE estimate of speech given this combination
  - Approach II
    - Find posterior distribution of models
    - Find MMSE estimate of speech given the posterior distribution
- Performance now sufficient for practical applications
  - Watch musical noise
  - Distortion versus noise suppression