The University Of Sheffield.

SPandH

Channel Compensation in the Generalised VTS Approach to Robust ASR

Erfan Loweimi
Jon Barker
Thomas Hain

Speech and Hearing Research Group (**SPandH**)

UKSpeech
2017

Outline

– **VTS** for Robust ASR

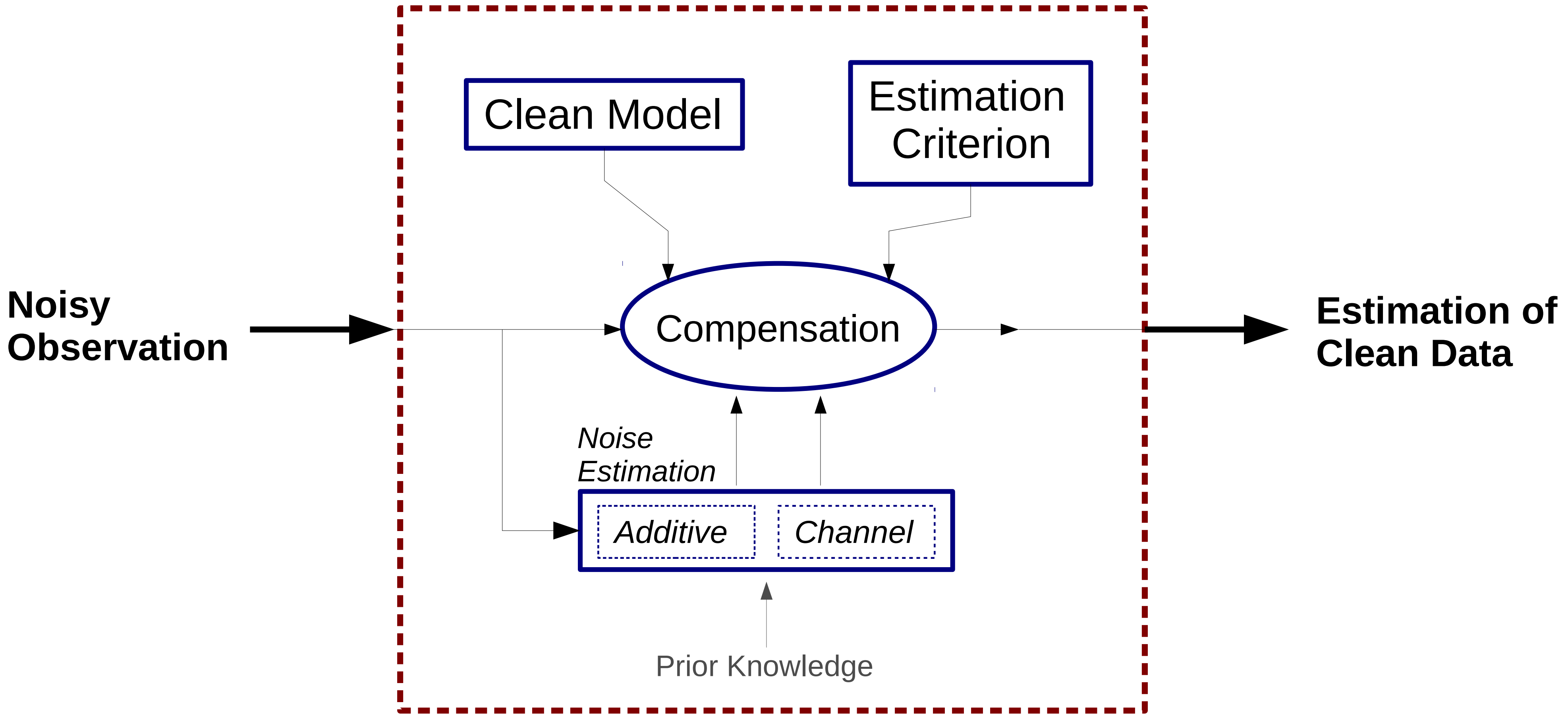
– **G**eneralised VTS

– Channel Noise Estimation

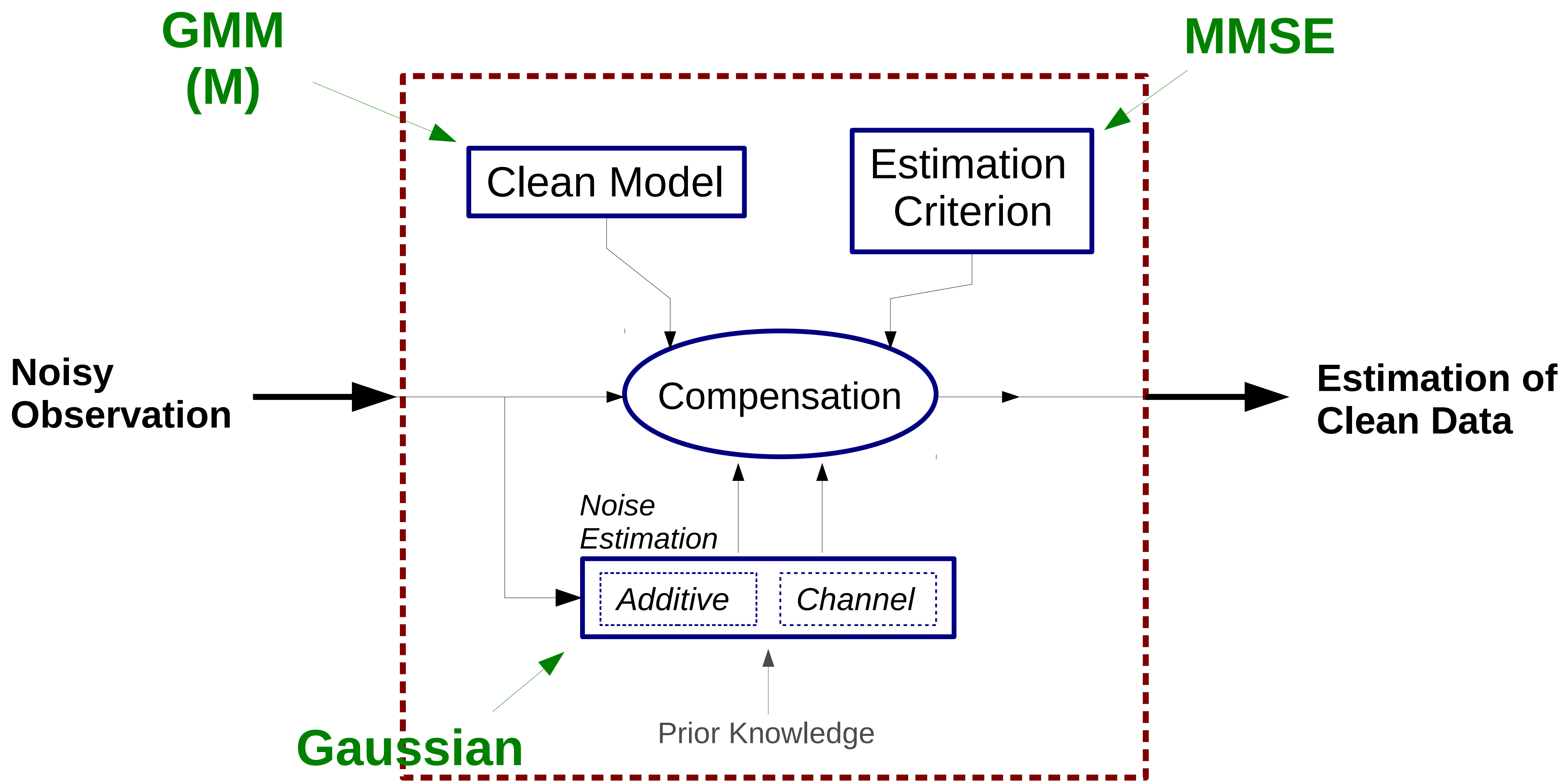
– Experimental Results

Vector Taylor Series (VTS) for Robust ASR

Model-based Noise Compensation

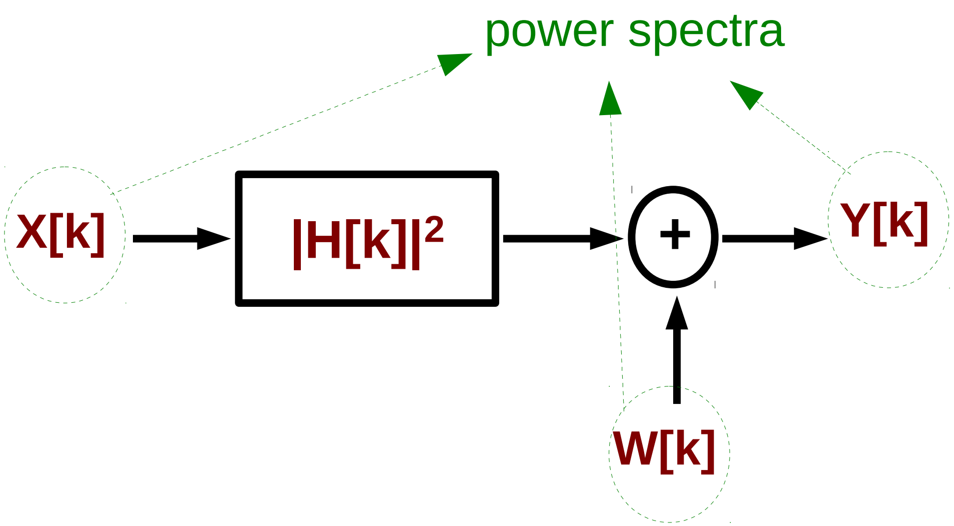


Model-based Noise Compensation



Environment Model

$$Y[k] = X[k]\,|H(k)|^2 + W[k]$$



$$\tilde{Y} = \tilde{X} + \underbrace{\tilde{H} + log\{1 + e^{\tilde{W} - \tilde{X} - \tilde{H}}\}}_{\tilde{G}(\tilde{X},\tilde{W},\tilde{H})}$$

$$\tilde{Z} = log Z$$

$$\tilde{Y} = \tilde{X} + \tilde{G}(\tilde{X},\tilde{W},\tilde{H})$$

Distortion function

Noise Compensation

$$\hat{X}_{MMSE} = \mathbb{E}[\tilde{X}|\tilde{Y}] = \int \tilde{X}\, p(\tilde{X}|\tilde{Y}) d\tilde{X}$$

Noise Compensation

$$\hat{X}_{MMSE} = \mathbb{E}[\tilde{X}|\tilde{Y}] = \int \tilde{X}\, p(\tilde{X}|\tilde{Y}) d\tilde{X}$$

$$\approx \tilde{Y} - \sum_{m=1}^{M} p(m|\tilde{Y})\ \tilde{G}(\mu_m^{\tilde{X}}, \mu^{\tilde{W}}, \mu^{\tilde{H}})$$

Assumptions ...

$$\hat{X}_{MMSE} = \mathbb{E}[\tilde{X}|\tilde{Y}] = \int \tilde{X}\, p(\tilde{X}|\tilde{Y}) d\tilde{X}$$

$$\approx \tilde{Y} - \sum_{m=1}^{M} p(m|\tilde{Y})\, \tilde{G}(\mu_m^{\tilde{X}}, \mu^{\tilde{W}}, \mu^{\tilde{H}})$$

- $\tilde{Y} \sim \sum_{m=1}^{M} p_{\tilde{Y}}(m)\, \mathcal{N}(\tilde{Y}; \mu_m^{\tilde{Y}}, \Sigma_m^{\tilde{Y}})$
- $\tilde{Y}$ and $\tilde{X}$ are jointly Gaussian within each component (Y?)
- $\tilde{G}$ is evaluated at the mean of the Gaussians (Y?)

Noise Compensation

$$\hat{X}_{MMSE} = \mathbb{E}[\tilde{X}|\tilde{Y}] = \int \tilde{X}\, p(\tilde{X}|\tilde{Y}) d\tilde{X}$$

$$\otimes \quad \approx \tilde{Y} - \sum_{m=1}^{M} p(m|\tilde{Y})\, \tilde{G}(\mu_m^{\tilde{X}}, \mu^{\tilde{W}}, \mu^{\tilde{H}})$$

$$\text{GMM of } \tilde{Y} : \ \theta^{\tilde{Y}} = \{p_m^{\tilde{Y}}, \mu_m^{\tilde{Y}}, \Sigma_m^{\tilde{Y}}\} \quad ???$$

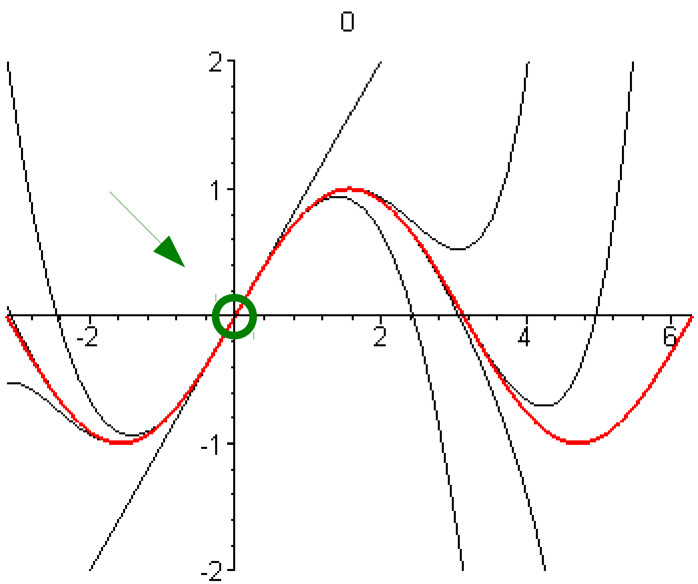
Non-linearity Problem

$$\tilde{Y} = \underbrace{\tilde{X} + \tilde{G}(\tilde{X}, \tilde{W}, \tilde{H})}_{f(\tilde{X}, \tilde{W}, \tilde{H})}$$

Non-linear

$$\text{GMM of } \tilde{Y}: \ \theta^{\tilde{Y}} = \{p_m^{\tilde{Y}}, \mu_m^{\tilde{Y}}, \Sigma_m^{\tilde{Y}}\} \quad ???$$

Taylor Series



$$\tilde{y} = f(\tilde{x}, \tilde{w}, \tilde{h})$$

Non-linear

Vector Taylor Series

$$\tilde{y} \approx \tilde{y}(\tilde{x}_0, \tilde{w}_0, \tilde{h}_0) + J^{\tilde{x}}(\tilde{x} - \tilde{x}_0) + J^{\tilde{w}}(\tilde{w} - \tilde{w}_0) + J^{\tilde{h}}(\tilde{h} - \tilde{h}_0)$$

$$J^{\tilde{z}}[i,j] = \left. \frac{\partial \tilde{y}_i}{\partial \tilde{z}_j} \right|_{(\tilde{\mathbf{w}}_\mathbf{0}, \tilde{\mathbf{x}}_\mathbf{0}, \tilde{\mathbf{h}}_\mathbf{0})} \quad \forall \quad z \in \{\tilde{x}, \tilde{w}, \tilde{h}\}$$

Requirements:

– Point(s)
– Jacobians ($J^{\tilde{z}}$)

VTS for ASR -- Points

$$\tilde{y} \approx \tilde{y}(\tilde{x}_0, \tilde{w}_0, \tilde{h}_0) + J^{\tilde{x}}(\tilde{x} - \tilde{x}_0) + J^{\tilde{w}}(\tilde{w} - \tilde{w}_0) + J^{\tilde{h}}(\tilde{h} - \tilde{h}_0)$$

$$J^{\tilde{z}}[i,j] = \left.\frac{\partial \tilde{y}_i}{\partial \tilde{z}_j}\right|_{(\mu^{\tilde{\mathbf{w}}}, \mu^{\tilde{\mathbf{x}}}_{\mathbf{m}}, \mu^{\tilde{\mathbf{h}}})} \quad \forall \quad z \in \{\tilde{x}, \tilde{w}, \tilde{h}\}$$

Requirements:

- Point(s) → **Means of the Gaussians**
- Jacobians ($J^{\tilde{z}}$)

VTS for ASR -- Jacobians

$$\tilde{y} \;\approx\; \tilde{y}(\tilde{x}_0, \tilde{w}_0, \tilde{h}_0) + J^{\tilde{x}}(\tilde{x} - \tilde{x}_0) + J^{\tilde{w}}\,(\tilde{w} - \tilde{w}_0) + J^{\tilde{h}}(\tilde{h} - \tilde{h}_0)$$

$$J^{\tilde{x}}_m = C\; diag\{\frac{1}{1 + V_m}\}\; C^{-1}$$

$$J^{\tilde{h}}_m = J^{\tilde{x}}_m = C\; diag\{\frac{1}{1 + V_m}\}\; C^{-1}$$

$$J^{\tilde{w}}_m = C\; C^{-1} - J^{\tilde{x}}_m = C\; diag\{\frac{V_m}{1 + V_m}\}\; C^{-1}$$

$$V_m = exp(C^{-1}(\mu^{\tilde{w}} - \mu^{\tilde{x}}_m - \mu^{\tilde{h}}))$$

VTS for ASR ...

$$\begin{cases} p_m^{\tilde{y}} & \approx p_m^{\tilde{x}} \\ \mu_m^{\tilde{y}} & \approx \mu_m^{\tilde{x}} + \mu^{\tilde{h}} + C\, log(1 + V_m) \\ \Sigma_m^{\tilde{y}} & \approx J_m^{\tilde{x}} \Sigma_m^{\tilde{x}} {J_m^{\tilde{x}}}^T + J_m^{\tilde{x}} \Sigma^{\tilde{h}} {J_m^{\tilde{h}}}^T + J_m^{\tilde{w}} \Sigma^{\tilde{w}} {J_m^{\tilde{w}}}^T \end{cases}$$

$$\hat{x}_{MMSE} \approx \tilde{y} - \sum_{m=1}^{M} p(m|\tilde{y})\ \tilde{g}(\mu^{\tilde{w}}, \mu_m^{\tilde{x}}, \mu^{\tilde{h}})$$

Generalised VTS

Generalised Non-linearity (GN)

$$\begin{cases} GenLog(x;\alpha) = \frac{1}{\alpha}(x^{\alpha} - 1), \quad x > 0 \quad \alpha \neq 0 \\ \lim_{\alpha \to 0} GenLog(x;\alpha) = log(x) \end{cases}$$

Generalised Non-linearity (GN)

- Statistics
  - Box-Cox Transformation (1964)
- Speech Processing
  - **Gen**eralised **Log**arithmic Function (1984)

$$\begin{cases} GenLog(x;\alpha) = \frac{1}{\alpha}(x^{\alpha} - 1), \quad x > 0 \quad \alpha \neq 0 \\ \lim_{\alpha \to 0} GenLog(x;\alpha) = log(x) \end{cases}$$

Advantages of GenLog

$$\begin{cases} GenLog(x;\alpha) = \frac{1}{\alpha}(x^{\alpha} - 1), \quad x > 0 \quad \alpha \neq 0 \\ \lim_{\alpha \to 0} GenLog(x;\alpha) = log(x) \end{cases}$$

[PDF] An Analysis of Transformations G. E. P. Box; D. R. Cox Journal of the…
https://pdfs.semanticscholar.org/6e82/0cf11712b9041bb625634612a535476f0960.pdf
by GEP Box - 1964 - Cited by 12876 - Related articles
29 Sep 2007 - **Box** AND **COX-An Analysis of Transformations**. [No. 2,. Each of the considerations (i)-(iii) can, and has been, used separately to select a.

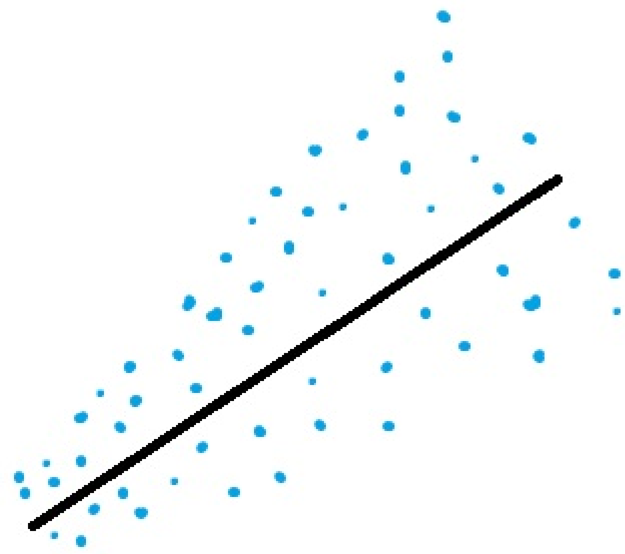
19, Aug., 2017

Advantages of GenLog

$$\begin{cases} GenLog(x;\alpha) = \frac{1}{\alpha}(x^{\alpha} - 1), \quad x > 0 \quad \alpha \neq 0 \\ \lim_{\alpha \to 0} GenLog(x;\alpha) = log(x) \end{cases}$$

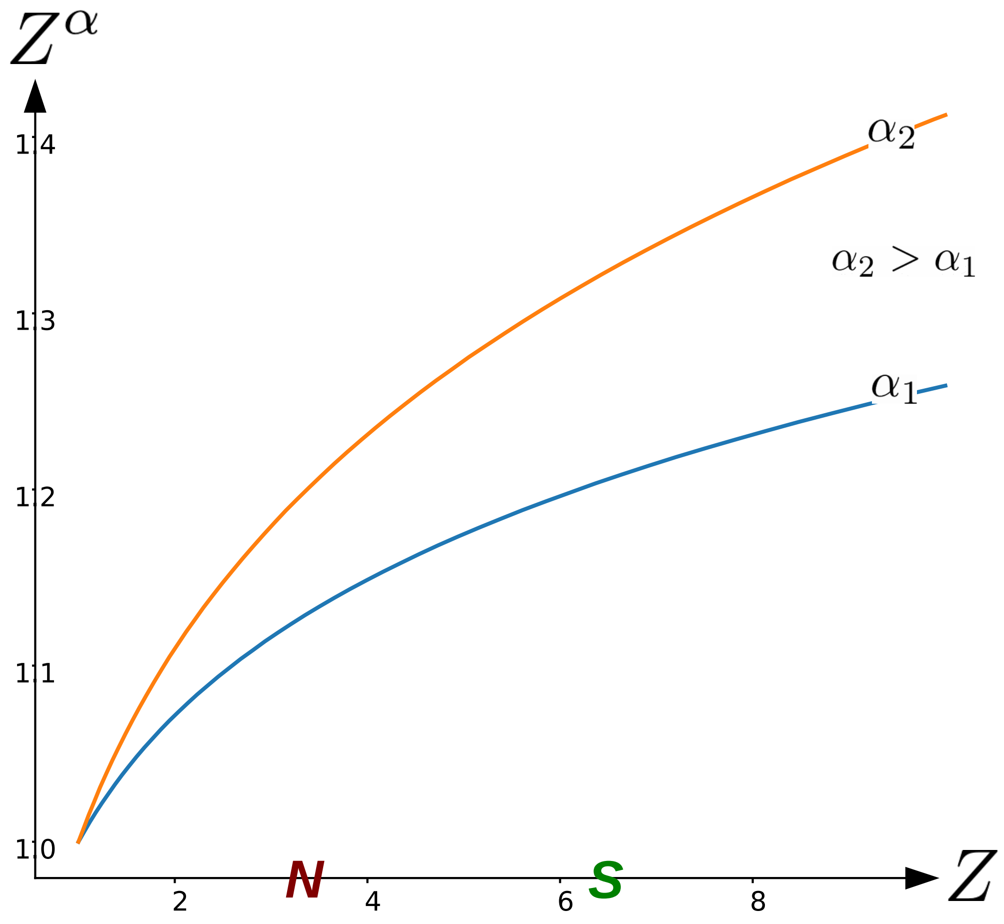
* Can potentially improve the ...

– *Linearity*
– *Homoscedasticity*
– ***Normality***



[PDF] An Analysis of Transformations G. E. P. Box; D. R. Cox Journal of the…
https://pdfs.semanticscholar.org/6e82/0cf11712b9041bb625634612a535476f0960.pdf
by GEP Box - 1964 - Cited by 12876 - Related articles
29 Sep 2007 - **Box** AND **COX-An Analysis of Transformations**. [No. 2,. Each of the considerations (i)-(iii) can, and has been, used separately to select a.

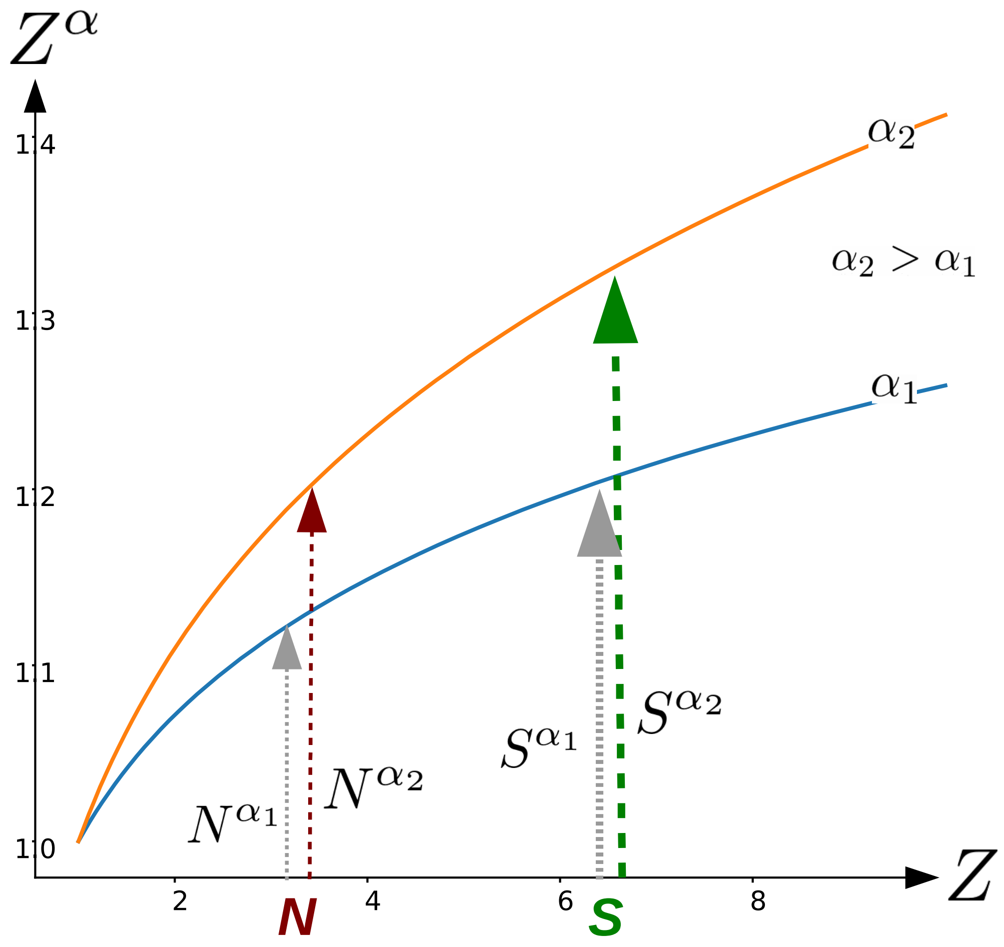
GenLog can improve the SNR!



GenLog can improve the SNR ...

$$SNR_1 = \frac{S^{\alpha_1}}{N^{\alpha_1}}$$

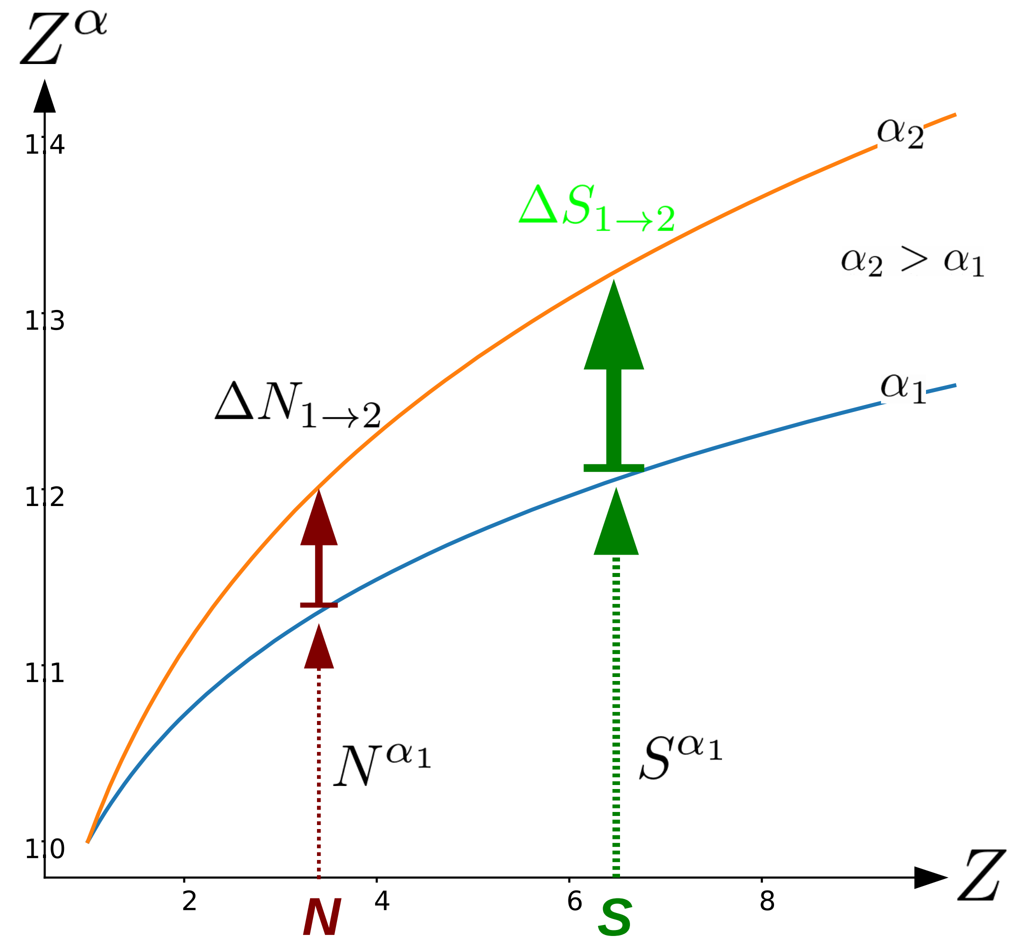
$$SNR_2 = \frac{S^{\alpha_2}}{N^{\alpha_2}}$$



GenLog can improve the SNR ...

$$SNR_1 = \frac{S^{\alpha_1}}{N^{\alpha_1}}$$

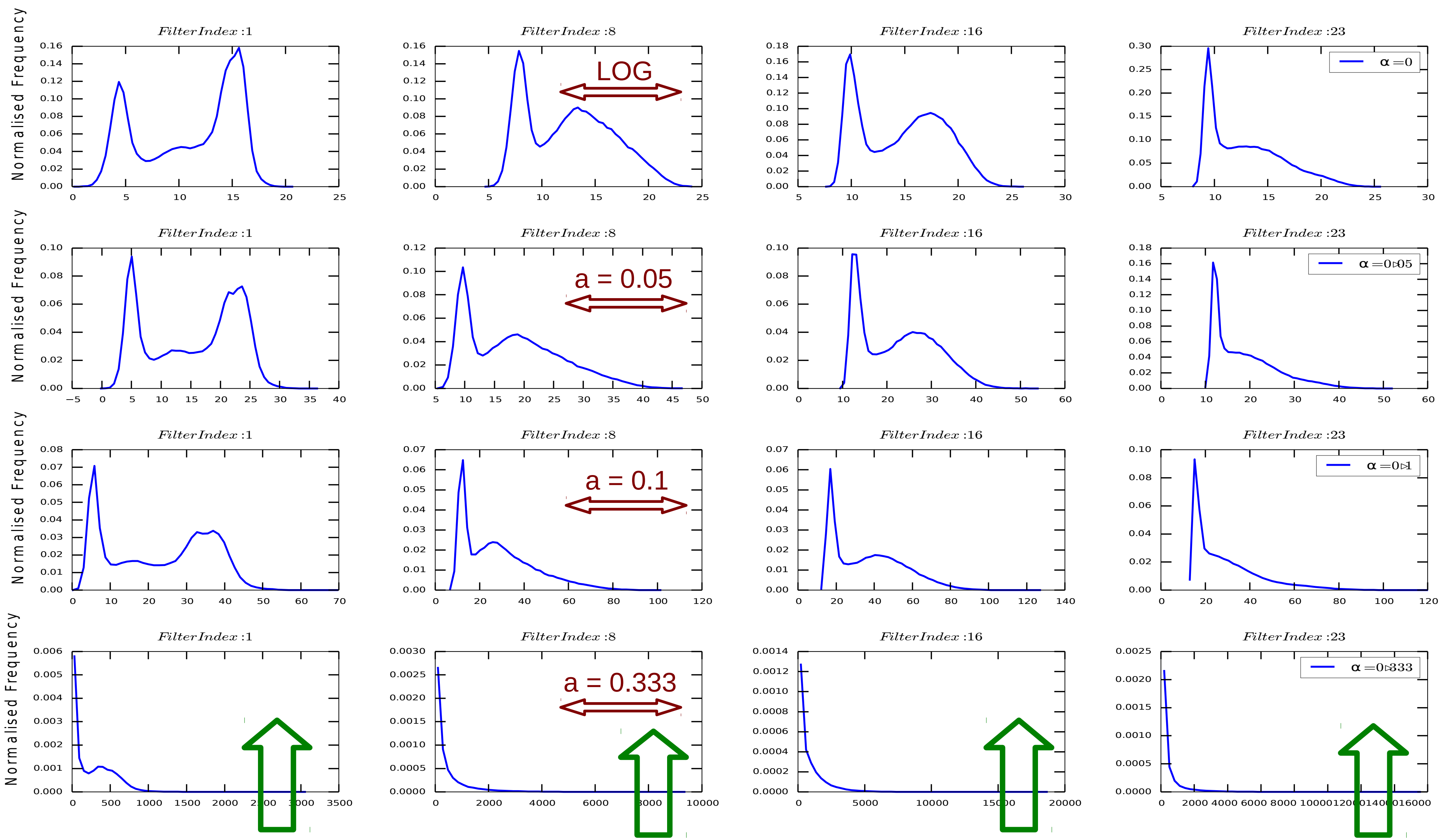
$$SNR_2 = \frac{S^{\alpha_2}}{N^{\alpha_2}} = \frac{S^{\alpha_1} + \Delta S_{1\to 2}}{N^{\alpha_1} + \Delta N_{1\to 2}}$$



$$\alpha_1 < \alpha_2 \Rightarrow SNR_1 < SNR_2$$

- NBins: 50
- 330 Utterances, WSJ
- #frames > 241 k

Statistical Effect of GenLog



Filter Index: 1 **Filter Index: 8** **Filter Index: 16** **Filter Index: 23**

$$
\begin{cases}
GenLog(x;\alpha) = \frac{1}{\alpha}(x^{\alpha} - 1), \quad x > 0 \quad \alpha \neq 0 \\
\lim_{\alpha \to 0} GenLog(x;\alpha) = log(x)
\end{cases}
$$

Statistical Distribution

SNR Boost

$$
\begin{cases} GenLog(x;\alpha) = \frac{1}{\alpha}(x^{\alpha} - 1), \quad x > 0 \quad \alpha \neq 0 \\ \lim_{\alpha \to 0} GenLog(x;\alpha) = log(x) \end{cases}
$$

$\downarrow \alpha$

Statistical Distribution

$\uparrow \alpha$

SNR Boost

$$\begin{cases} GenLog(x;\alpha) = \frac{1}{\alpha}(x^{\alpha} - 1), \quad x > 0 \quad \alpha \neq 0 \\ \lim_{\alpha \to 0} GenLog(x;\alpha) = log(x) \end{cases}$$

$\downarrow \alpha$

$\uparrow \alpha$

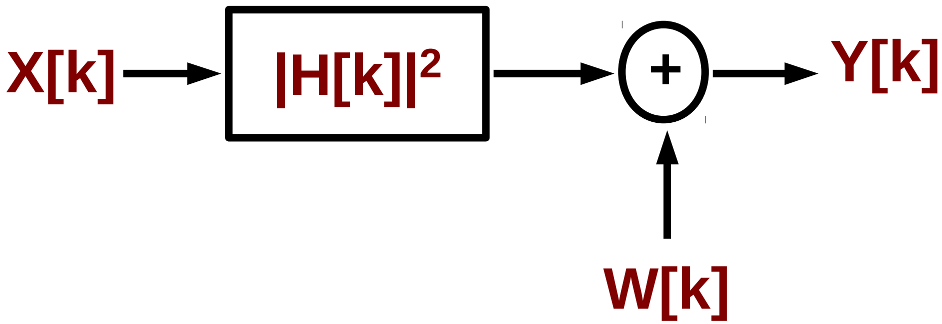
Statistical Distribution

SNR Boost

$$0.05 \leq \alpha < 0.1$$

Environment Model

$$Y[k] = X[k]\,|H(k)|^2 + W[k]$$



$$\breve{Y} \;=\; \breve{X}\;\underbrace{\breve{H}\,(1+(\frac{\breve{W}}{\breve{X}\breve{H}})^{\frac{1}{\alpha}})^{\alpha}}_{\breve{G}(\breve{X},\breve{W},\breve{H})}$$

$$\breve{Z} = Z^{\alpha}$$

$$\breve{Y} \;=\; \breve{X}\,\breve{G}(\breve{X},\breve{W},\breve{H})$$

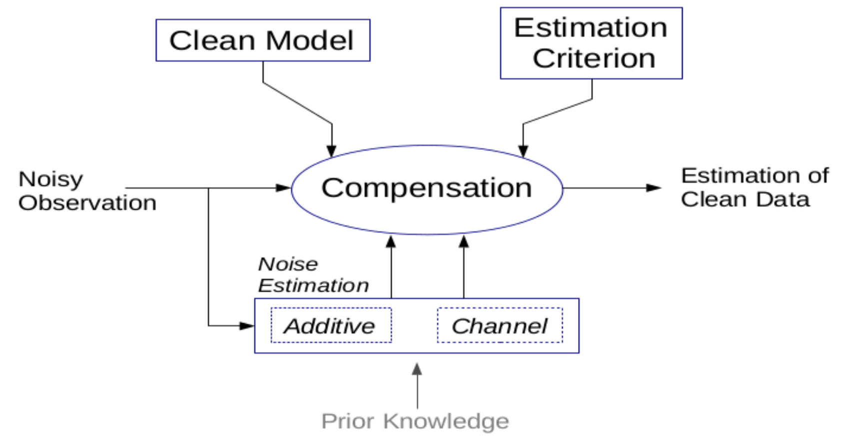
Distortion function

Generalised VTS (gVTS)

$$\breve{X}_{MMSE} = \mathbb{E}[\breve{X}|\breve{Y}] = \int \breve{X}\, p(\breve{X}|\breve{Y}) d\breve{X}$$

$$\approx \breve{Y} \sum_{m=1}^{M} p(m|\breve{Y}) \frac{1}{\breve{G}(\mu_m^{\breve{X}}, \mu^{\breve{W}}, \mu^{\breve{H}})}$$

gVTS → Equations



$$Y[k] = X[k]\,|H(k)|^2 + W[k]$$

$$J^{\lambda}_{gVTS} = \frac{\partial \breve{Y}}{\partial \lambda} = \frac{\partial \breve{Y}}{\partial Y}\frac{\partial Y}{\partial \lambda} = 2\alpha \frac{\mu^{\breve{X}}_m \mu^{\breve{H}} \sqrt{V_m}}{(1+V_m)^{1-\alpha}}$$

$$\begin{cases} GenLog(z;\alpha) = \frac{1}{\alpha}(z^{\alpha}-1), \quad z>0 \quad \alpha \neq 0 \\ \lim_{\alpha\to 0} GenLog(z;\alpha) = log(z), \end{cases}$$

$$\Sigma^{\breve{Y}}_m \leftarrow \Sigma^{\breve{Y}}_m + J^{\lambda}_{VTS}\,\Sigma^{\lambda}\,{J^{\lambda}_{VTS}}^T$$

$$J^{\breve{X}}_m = \left.\frac{\partial \breve{Y}}{\partial \breve{X}}\right|_{(\mu^{\breve{X}}_m, \mu^{\breve{H}}, \mu^{\breve{W}})} = diag\{\mu^{\breve{H}}\ (1+\breve{V}_m)^{\alpha-1}\}$$

$$\breve{Y} = \breve{X}\,\breve{H}\,(1 + (\frac{\breve{W}}{\breve{X}\breve{H}})^{\frac{1}{\alpha}})^{\alpha}$$

$$J^{\breve{H}}_m = \left.\frac{\partial \breve{Y}}{\partial \breve{H}}\right|_{(\mu^{\breve{X}}_m, \mu^{\breve{H}}, \mu^{\breve{W}})} = diag\{\mu^{\breve{X}}_m\ (1+\breve{V}_m)^{\alpha-1}\}$$

$$G(\breve{X},\breve{H},\breve{W}) = (1 + (\frac{\breve{W}}{\breve{X}\breve{H}})^{\frac{1}{\alpha}})^{\alpha}$$

$$J^{\breve{W}}_m = \left.\frac{\partial \breve{Y}}{\partial \breve{W}}\right|_{(\mu^{\breve{X}}_m, \mu^{\breve{H}}, \mu^{\breve{W}})} = diag\{\mu^{\breve{H}}\ (\frac{1+\breve{V}_m}{\breve{V}_m})^{\alpha-1}\}$$

$$\begin{pmatrix}\tilde{X}\\ \tilde{Y}\end{pmatrix}_m \sim \mathcal{N}\left[\begin{pmatrix}\mu^{\tilde{X}}_m\\ \mu^{\tilde{Y}}_m\end{pmatrix}, \begin{pmatrix}\Sigma_{\tilde{X}\tilde{X}} & \Sigma_{\tilde{X}\tilde{Y}}\\ \Sigma_{\tilde{Y}\tilde{X}} & \Sigma_{\tilde{Y}\tilde{Y}}\end{pmatrix}\right]$$

$$\begin{cases} \breve{X} \sim \sum_{m=1}^{M} p_{\breve{x}}(m)\,\mathcal{N}(\mu^{\breve{X}}_m, \Sigma^{\breve{X}}_m) \\ \breve{W} \sim \mathcal{N}(\mu^{\breve{W}}, \Sigma^{\breve{W}}), \\ \breve{H} \sim \mathcal{N}(\mu^{\breve{H}}, \Sigma^{\breve{H}}), \end{cases}$$

$$\breve{V}_m = (\frac{\mu^{\breve{W}}}{\mu^{\breve{X}}_m\,\mu^{\breve{H}}})^{\frac{1}{\alpha}}$$

$$\mathcal{E}\{\frac{1}{\breve{X}_u}\} \geq \frac{1}{\mathcal{E}\{\breve{X}_u\}}$$

$$J^{\tilde{z}} = \frac{\partial \tilde{\mathbf{y}}}{\partial \tilde{\mathbf{z}}} = \frac{\partial\,\mathbf{C}\tilde{\mathbf{Y}}}{\partial\,\mathbf{C}\tilde{\mathbf{Z}}} = C\,\frac{\partial \tilde{\mathbf{Y}}}{\partial \tilde{\mathbf{Z}}}\,C^{-1} \Rightarrow J^{\tilde{z}} = C\,J^{\tilde{Z}}\,C^{-1}$$

$$\breve{X}_{MMSE} = \mathcal{E}[\breve{X}|\breve{Y}] = \int \breve{X}\,p(\breve{X}|\breve{Y})d\breve{X}$$

$$\mu^{\breve{Y}}_m \approx \mu^{\breve{X}}_m\,\mu^{\breve{H}}(1 + (\frac{\mu^{\breve{W}}}{\mu^{\breve{X}}_m\mu^{\breve{H}}})^{\frac{1}{\alpha}})^{\alpha}$$

$$\breve{X}_{MMSE} = \int \frac{\breve{Y}}{G(\breve{X},\breve{H},\breve{W})} \sum_{m=1}^{M} p(\breve{X}|m)\,p(m|\breve{Y})\,d\breve{X}$$
$$= \breve{Y} \sum_{m=1}^{M} p(m|\breve{Y})\,\frac{1}{G(\mu^{\breve{X}}_m, \mu^{\breve{H}}, \mu^{\breve{W}})}$$

$$\Sigma^{\breve{Y}}_m \approx J^{\breve{X}}_m \Sigma^{\breve{X}}_m\,{J^{\breve{X}}_m}^T + J^{\breve{W}}_m\,\Sigma^{\breve{W}}\,{J^{\breve{W}}_m}^T + J^{\breve{H}}_m\,\Sigma^{\breve{h}}\,{J^{\breve{H}}_m}^T$$

$$\mathcal{E}\{\breve{X}_u\} \approx \sum_{m=1}^{M} p_x(m)\,\mu^{\breve{X}}_m$$

$$\breve{x}[n,k] = \frac{\breve{X}[n,k]\,|H[k]|^{2\gamma}}{\sqrt[N]{\prod_{n=1}^{N}\breve{X}[n,k]\,|H[k]|^{2\gamma}}}$$

$$Z \sim \sum_{m=1}^{M} p_x(m)\,\mathcal{N}(z; H_d\,\mu^{\breve{X}}_m, H_d\,\Sigma^{\breve{X}}_m\,H_d^T)$$

$$p(m|\breve{Y}) = \frac{p_{\breve{y}}(m)\,\mathcal{N}(\mu^{\breve{Y}}_m, \Sigma^{\breve{Y}}_m)}{\sum_{m'=1}^{M} p_{\breve{y}}(m')\,\mathcal{N}(\mu^{\breve{Y}}_{m'}, \Sigma^{\breve{Y}}_{m'})}$$

$$\breve{H}_t = \frac{\breve{Y}_t}{\breve{X}_t} \Rightarrow \mu^{\breve{H}} = \mathcal{E}\{\breve{H}\} = \mathcal{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\}$$

$$\mu^{\breve{H}} \approx \frac{\frac{1}{T}\sum_{t=1}^{T}\breve{Y}_t}{\sum_{m=1}^{M} p_x(m)\mu^{\breve{X}}_m}$$

$$\mu^{\breve{H}} = \mathcal{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\} = \mathcal{E}\{\breve{Y}_u\}\,\mathcal{E}\{\frac{1}{\breve{X}}\}$$

$$\frac{\mathcal{E}\{\breve{Y}_u\}}{\mathcal{E}\{\breve{X}_u\}} = \mathcal{E}\{(H + \frac{W_u}{X_u})^{\alpha}\} \approx \mu^{\breve{H}} + \mathcal{E}\{\frac{\breve{W}_u}{\breve{X}_u}\}$$

$$\breve{\mathbf{Y}} \approx \breve{Y}(\breve{X}_0, \breve{W}_0, \breve{H}_0) + J^{\breve{X}}(\breve{X} - \breve{X}_0) + J^{\breve{W}}(\breve{W} - \breve{W}_0) + J^{\breve{H}}(\breve{H} - \breve{H}_0)$$

$$\mathcal{E}\{\breve{Y}_u\} \approx \frac{1}{T}\sum_{t=1}^{T}\breve{Y}_t$$

$$\breve{\mathbf{Y}} = C^{-1}\,\breve{\mathbf{y}} \Rightarrow \begin{cases} p_{\breve{\mathbf{Y}}}(m) = p_{\breve{\mathbf{y}}}(m) \\ p(\breve{\mathbf{Y}}|m) = p(\breve{\mathbf{y}}|m) \end{cases} \Rightarrow p(m|\breve{\mathbf{Y}}) = p(m|\breve{\mathbf{y}})$$

SPandH

Channel Noise Estimation

Channel Noise Estimation

$$\breve{H}_t = \frac{\breve{Y}_t}{\breve{X}_t} \Rightarrow \mu^{\breve{H}} = \mathbb{E}\{\breve{H}\} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\}$$

Channel Noise Estimation

$$\breve{H}_t = \frac{\breve{Y}_t}{\breve{X}_t} \Rightarrow \mu^{\breve{H}} = \mathbb{E}\{\breve{H}\} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\}$$

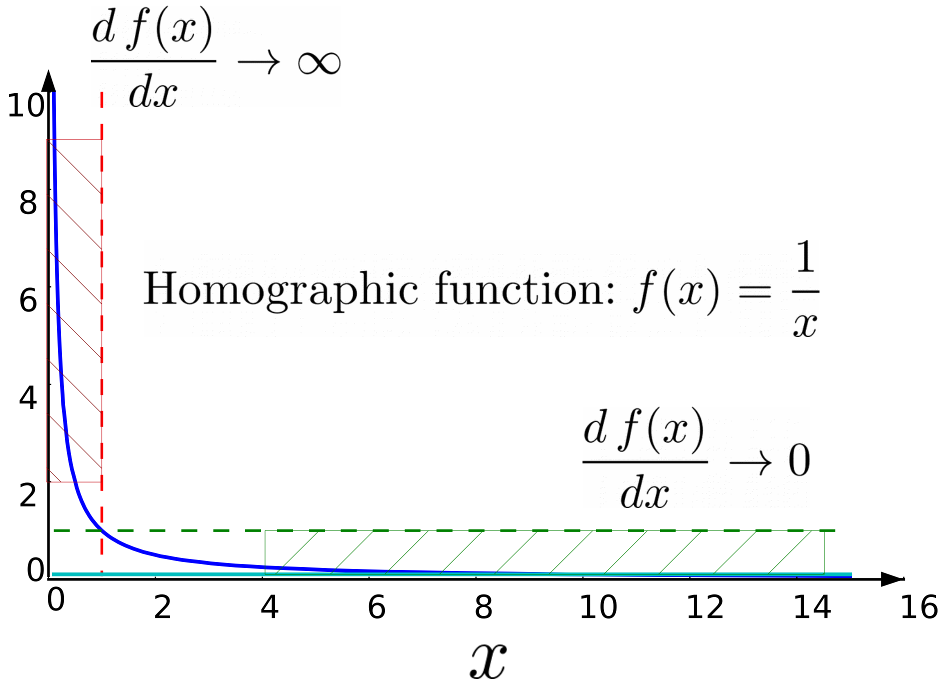
$$\mu^{\breve{H}} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\} \approx \mathbb{E}\{\breve{Y}_u\}\,\mathbb{E}\{\frac{1}{\breve{X}_u}\}$$

Y and $\frac{1}{X}$ are uncorrelated!

Channel Noise Estimation

$$\breve{H}_t = \frac{\breve{Y}_t}{\breve{X}_t} \Rightarrow \mu^{\breve{H}} = \mathbb{E}\{\breve{H}\} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\}$$

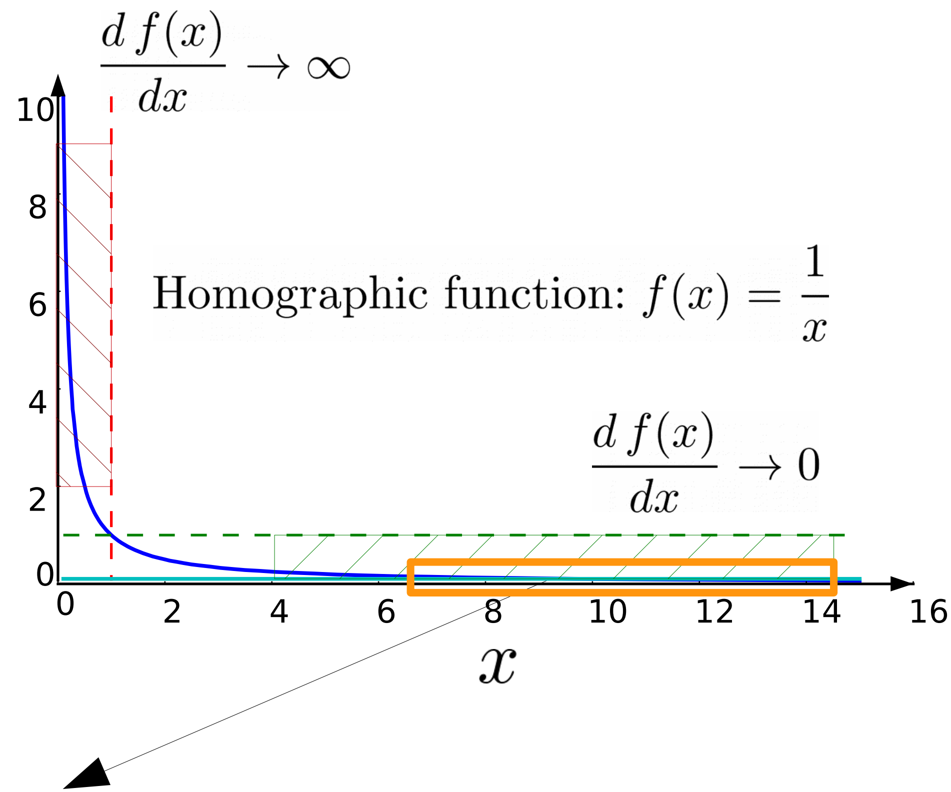
$$\mu^{\breve{H}} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\} \approx \mathbb{E}\{\breve{Y}_u\}\,\mathbb{E}\{\frac{1}{\breve{X}_u}\}$$



Channel Noise Estimation

$$\breve{H}_t = \frac{\breve{Y}_t}{\breve{X}_t} \Rightarrow \mu^{\breve{H}} = \mathbb{E}\{\breve{H}\} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\}$$

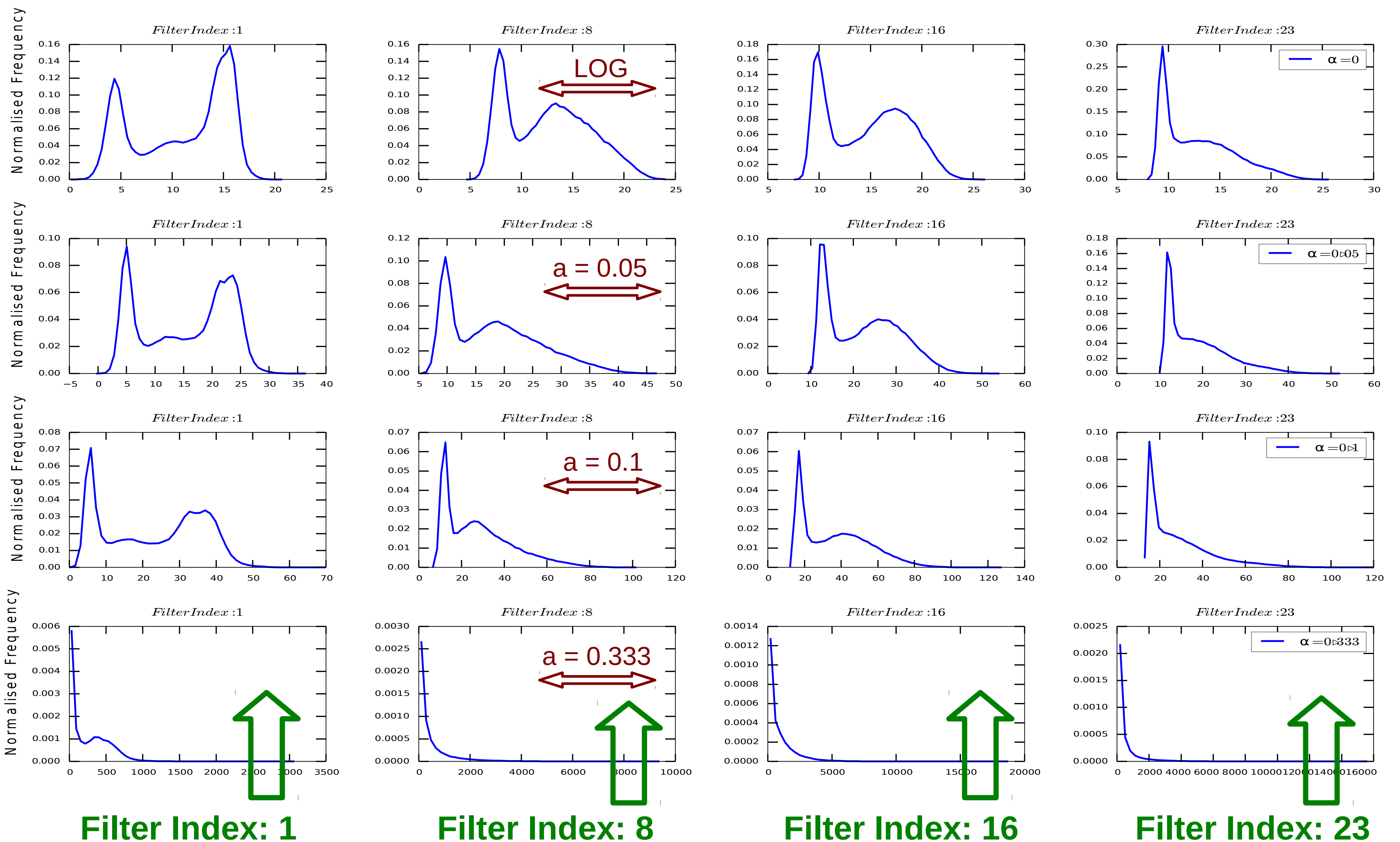
$$\mu^{\breve{H}} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\} \approx \mathbb{E}\{\breve{Y}_u\}\,\mathbb{E}\{\frac{1}{\breve{X}_u}\}$$



$\frac{1}{x}$ does not covary with x !

- NBins: 50
- 330 Utterances, WSJ
- #frames > 241 k

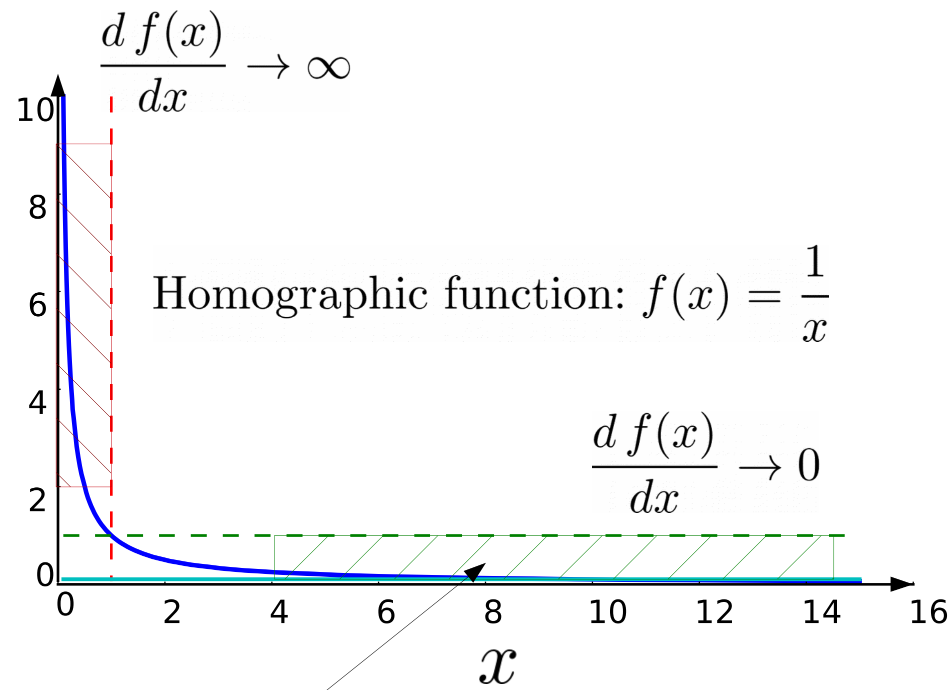
Statistical Effect of GenLog



Channel Noise Estimation

$$\breve{H}_t = \frac{\breve{Y}_t}{\breve{X}_t} \Rightarrow \mu^{\breve{H}} = \mathbb{E}\{\breve{H}\} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\}$$

$$\mu^{\breve{H}} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\} \approx \mathbb{E}\{\breve{Y}_u\}\,\mathbb{E}\{\frac{1}{\breve{X}_u}\}$$



x and $\frac{1}{x}$ are *almost* uncorreleted **here**!

Channel Noise Estimation

$$\breve{H}_t = \frac{\breve{Y}_t}{\breve{X}_t} \Rightarrow \mu^{\breve{H}} = \mathbb{E}\{\breve{H}\} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\}$$

STEP 1

$$\mu^{\breve{H}} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\} \approx \mathbb{E}\{\breve{Y}_u\}\,\mathbb{E}\{\frac{1}{\breve{X}_u}\}$$

Channel Noise Estimation

$$\breve{H}_t = \frac{\breve{Y}_t}{\breve{X}_t} \Rightarrow \mu^{\breve{H}} = \mathbb{E}\{\breve{H}\} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\}$$

STEP 1

$$\mu^{\breve{H}} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\} \approx \mathbb{E}\{\breve{Y}_u\}\ \mathbb{E}\{\frac{1}{\breve{X}_u}\}$$

2

$$\mathbb{E}\{\breve{Y}_u\} \approx \frac{1}{T}\sum_{t=1}^{T}\breve{Y}_t$$

2

$$\mathbb{E}\{\breve{X}_u\} \approx \sum_{m=1}^{M} p_{\breve{x}}(m)\ \mu_m^{\breve{X}}$$

Channel Noise Estimation

$$\breve{H}_t = \frac{\breve{Y}_t}{\breve{X}_t} \Rightarrow \mu^{\breve{H}} = \mathbb{E}\{\breve{H}\} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\}$$

STEP 1

$$\mu^{\breve{H}} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\} \approx \mathbb{E}\{\breve{Y}_u\}\,\mathbb{E}\{\frac{1}{\breve{X}_u}\}$$

2

$$\mathbb{E}\{\breve{Y}_u\} \approx \frac{1}{T}\sum_{t=1}^{T} \breve{Y}_t$$

2

$$\mathbb{E}\{\breve{X}_u\} \approx \sum_{m=1}^{M} p_{\breve{x}}(m)\ \mu_m^{\breve{X}}$$

3

3

$$\mathbb{E}\{\frac{1}{\breve{X}_u}\} \geq \frac{1}{\mathbb{E}\{\breve{X}_u\}}$$

$\frac{1}{X}$ is convex !

Channel Noise Estimation

$$\breve{H}_t = \frac{\breve{Y}_t}{\breve{X}_t} \Rightarrow \mu^{\breve{H}} = \mathbb{E}\{\breve{H}\} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\}$$

STEP 1

$$\mu^{\breve{H}} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\} \approx \mathbb{E}\{\breve{Y}_u\}\, \mathbb{E}\{\frac{1}{\breve{X}_u}\}$$

2

$$\mathbb{E}\{\breve{Y}_u\} \approx \frac{1}{T}\sum_{t=1}^{T} \breve{Y}_t$$

2

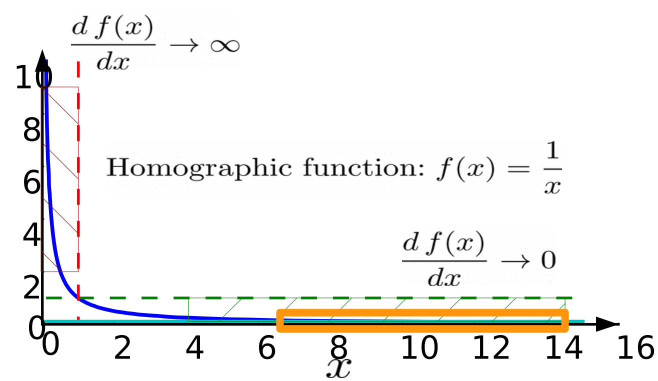
$$\mathbb{E}\{\breve{X}_u\} \approx \sum_{m=1}^{M} p_{\breve{x}}(m)\, \mu_m^{\breve{X}}$$

3

3

$$\mathbb{E}\{\frac{1}{\breve{X}_u}\} \approx \frac{1}{\mathbb{E}\{\breve{X}_u\}}$$

@ the border of
Convex/Concave



Channel Noise Estimation

$$\breve{H}_t = \frac{\breve{Y}_t}{\breve{X}_t} \Rightarrow \mu^{\breve{H}} = \mathbb{E}\{\breve{H}\} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\}$$

STEP 1

$$\mu^{\breve{H}} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\} \approx \mathbb{E}\{\breve{Y}_u\}\, \mathbb{E}\{\frac{1}{\breve{X}_u}\}$$

2

$$\mathbb{E}\{\breve{Y}_u\} \approx \frac{1}{T}\sum_{t=1}^{T} \breve{Y}_t$$

2

$$\mathbb{E}\{\breve{X}_u\} \approx \sum_{m=1}^{M} p_{\breve{x}}(m)\, \mu_m^{\breve{X}}$$

3

3

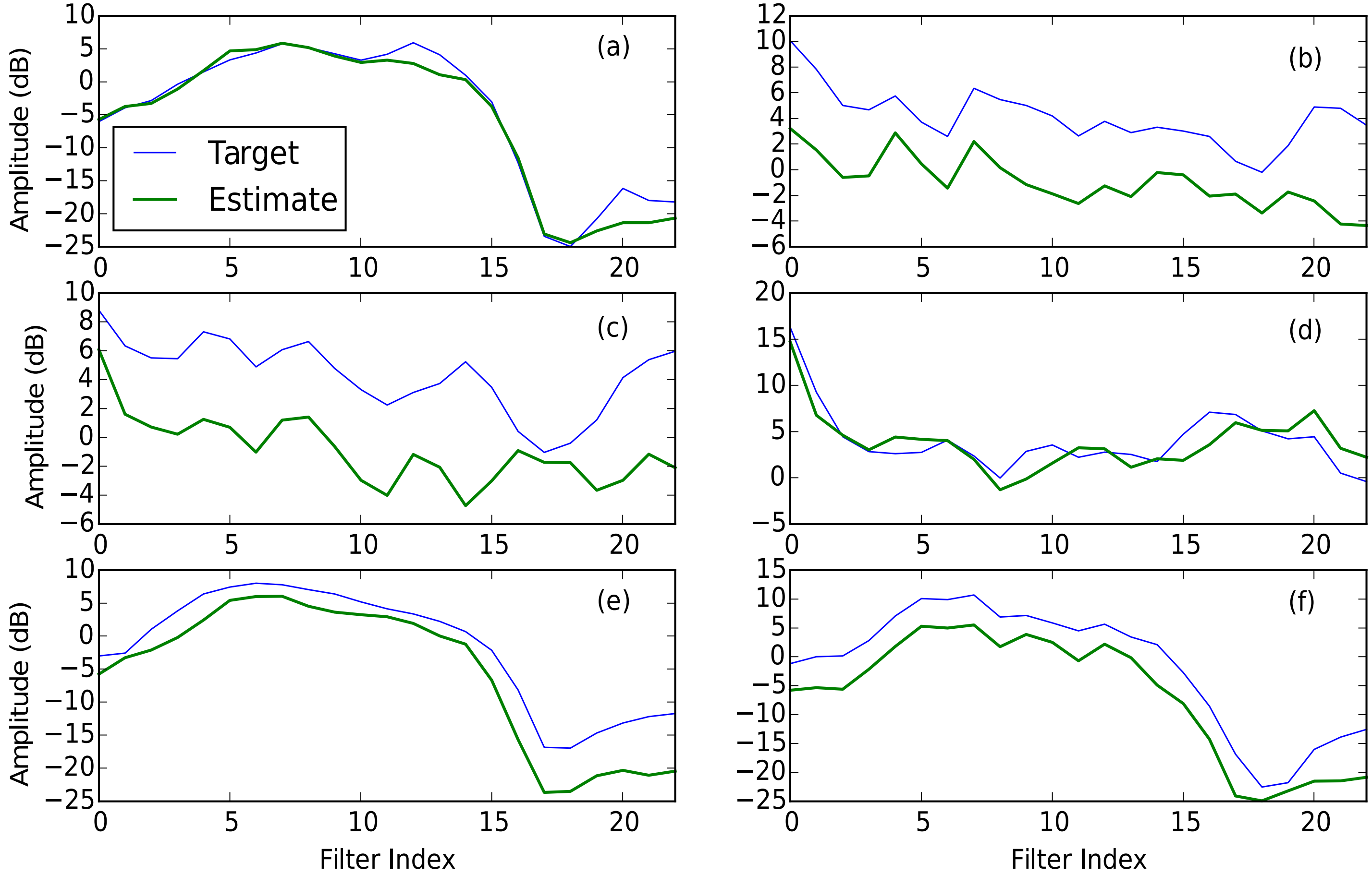
$$\mathbb{E}\{\frac{1}{\breve{X}_u}\} \approx \frac{1}{\mathbb{E}\{\breve{X}_u\}}$$

4

$$\mu^{\breve{H}} \approx \frac{\frac{1}{T}\sum_{t=1}^{T} \breve{Y}_t}{\sum_{m=1}^{M} p_{\breve{x}}(m)\, \mu_m^{\breve{X}}}$$

! Under Estimation

Channel Estimation
Aurora-4



$$\mu^{\breve{H}} = \mathbb{E}\{\breve{H}\} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\}$$

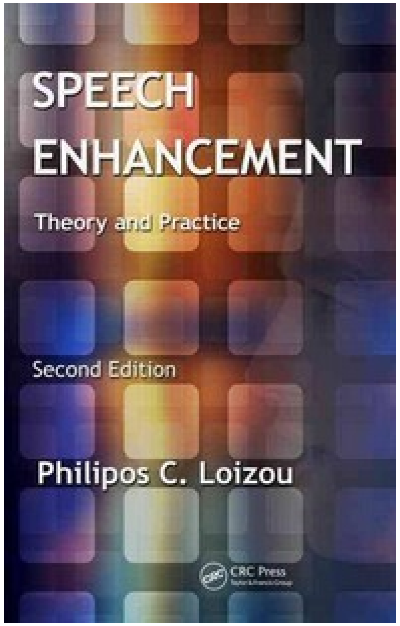
Y = XH ← Test Set C
X ← Test Set A

Additive Noise Effect

$$\mathbb{E}\{\frac{\breve{Y}}{\breve{X}}\} = \mathbb{E}\{\left(\frac{XH + W}{X}\right)^{\alpha}\} \approx \mu^{\breve{H}} + \mathbb{E}\{\frac{\breve{W}}{\breve{X}}\}$$

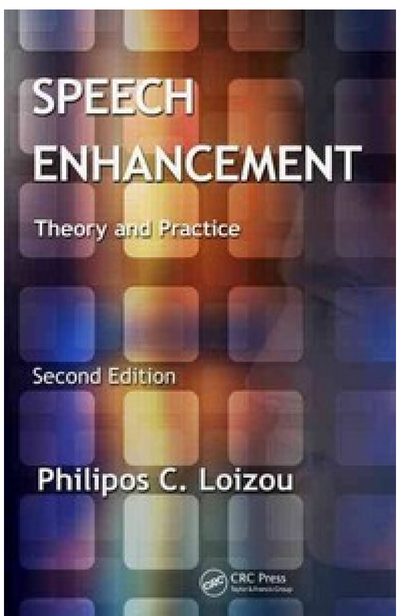
Additive Noise Effect

$$\mathbb{E}\{\frac{\breve{Y}}{\breve{X}}\} = \mathbb{E}\{(\frac{XH+W}{X})^{\alpha}\} \approx \mu^{\breve{H}} + \mathbb{E}\{\frac{\breve{W}}{\breve{X}}\}$$

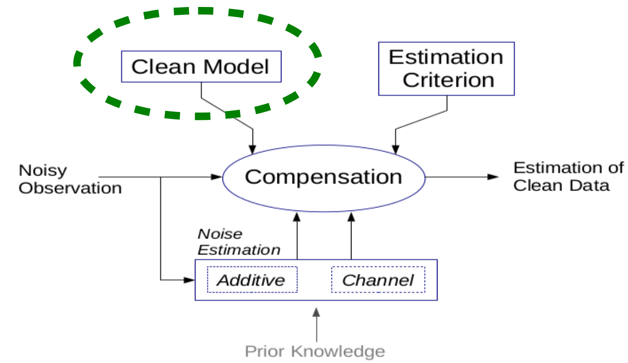


Additive Noise Effect

$$\mathbb{E}\{\frac{\breve{Y}}{\breve{X}}\} = \mathbb{E}\{(\frac{XH+W}{X})^{\alpha}\} \approx \mu^{\breve{H}} + \boxed{\mathbb{E}\{\frac{\breve{W}}{\breve{X}}\}}$$

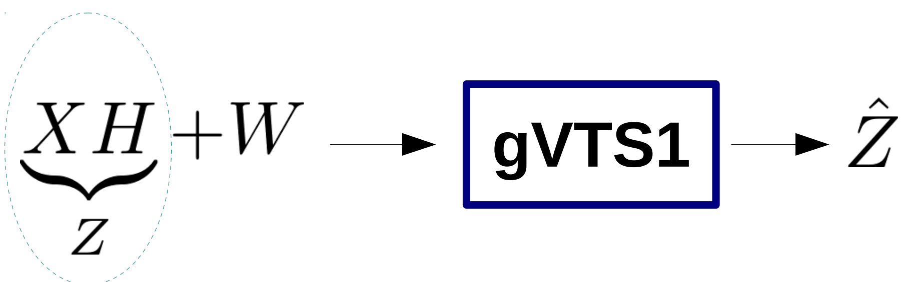


GMM



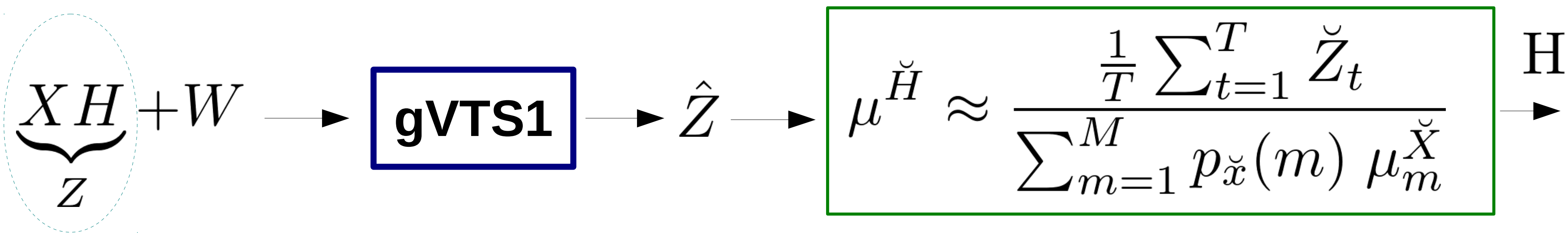
Additive Noise Effect

$$\mathbb{E}\{\frac{\breve{Y}}{\breve{X}}\} = \mathbb{E}\{(\frac{XH+W}{X})^{\alpha}\} \approx \mu^{\breve{H}} + \mathbb{E}\{\frac{\breve{W}}{\breve{X}}\}$$



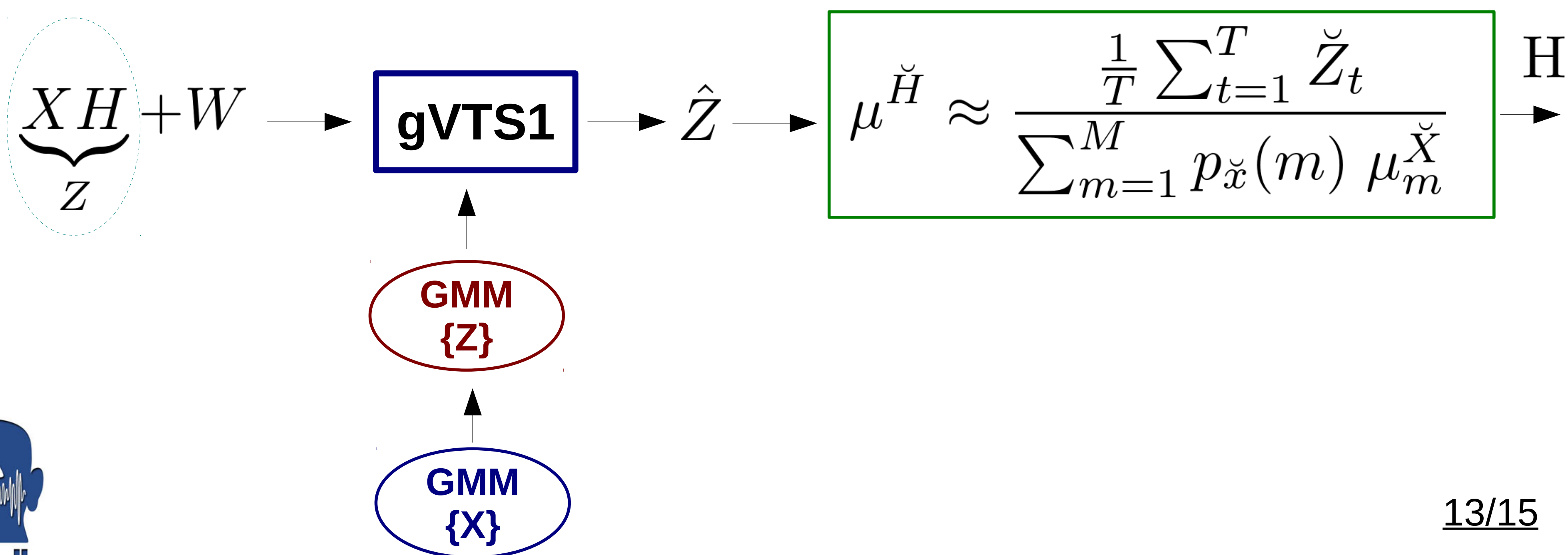
Additive Noise Effect

$$\mathbb{E}\{\frac{\breve{Y}}{\breve{X}}\} = \mathbb{E}\{(\frac{XH+W}{X})^{\alpha}\} \approx \mu^{\breve{H}} + \mathbb{E}\{\frac{\breve{W}}{\breve{X}}\}$$



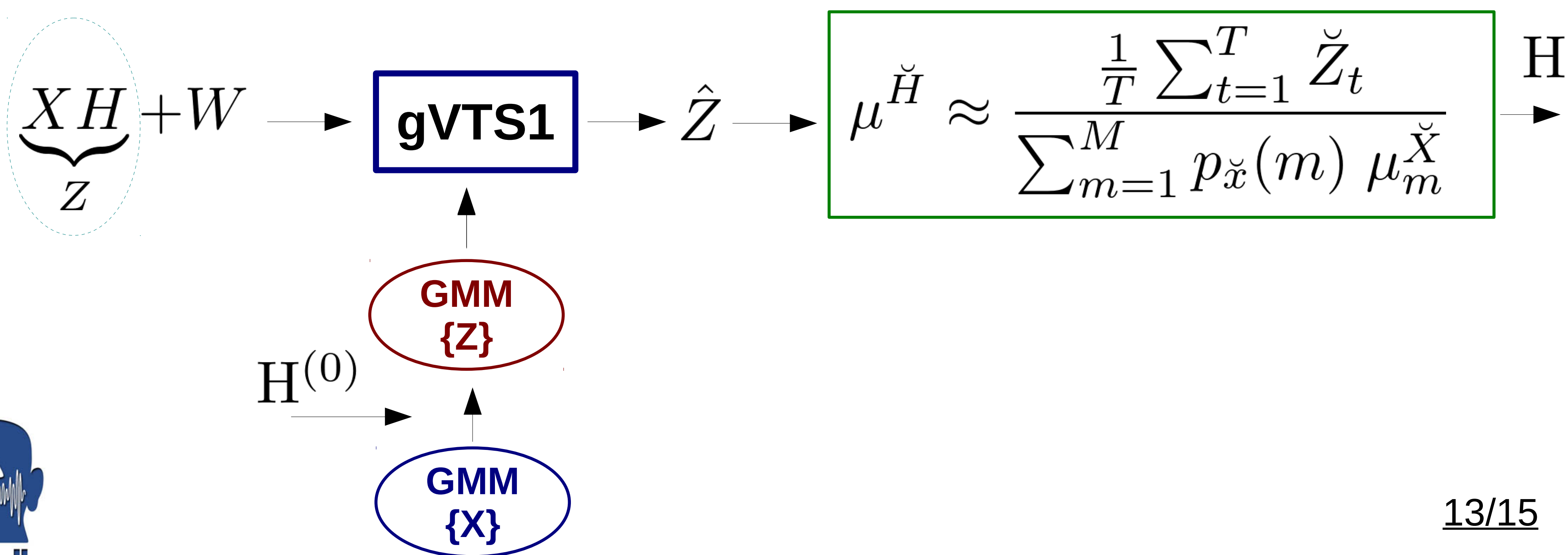
Additive Noise Effect

$$\mathbb{E}\{\frac{\breve{Y}}{\breve{X}}\} = \mathbb{E}\{(\frac{XH+W}{X})^{\alpha}\} \approx \mu^{\breve{H}} + \mathbb{E}\{\frac{\breve{W}}{\breve{X}}\}$$



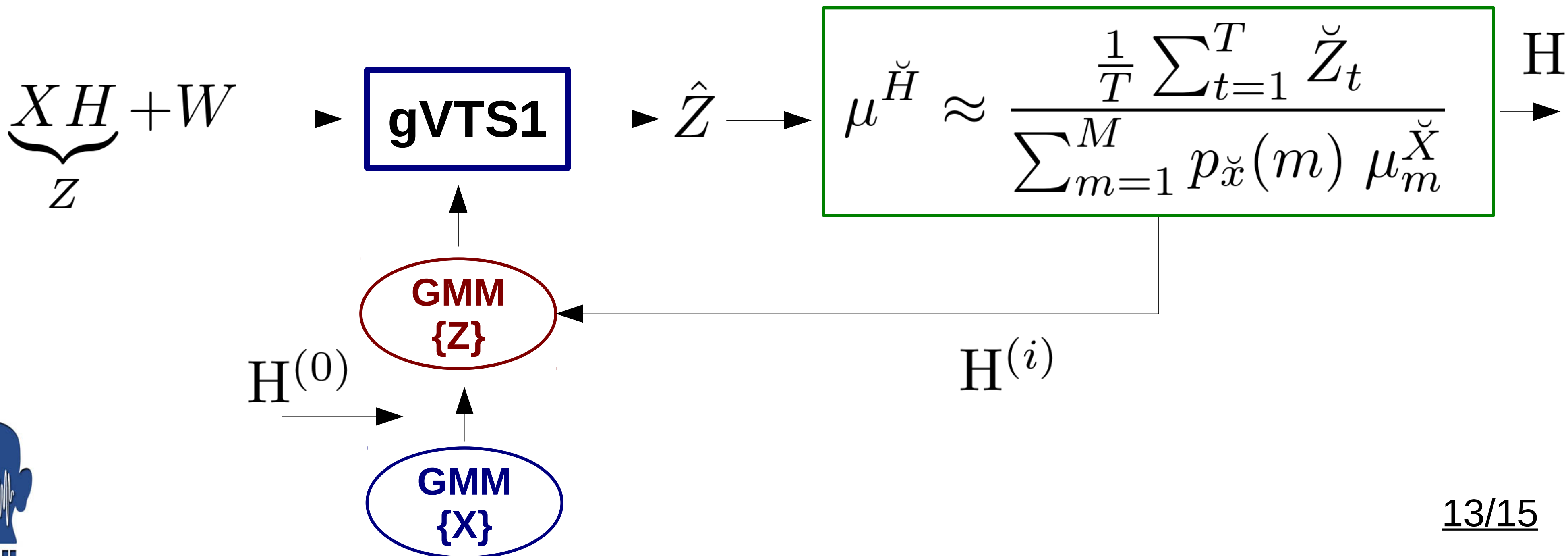
Additive Noise Effect

$$\mathbb{E}\{\frac{\breve{Y}}{\breve{X}}\} = \mathbb{E}\{(\frac{XH+W}{X})^{\alpha}\} \approx \mu^{\breve{H}} + \mathbb{E}\{\frac{\breve{W}}{\breve{X}}\}$$



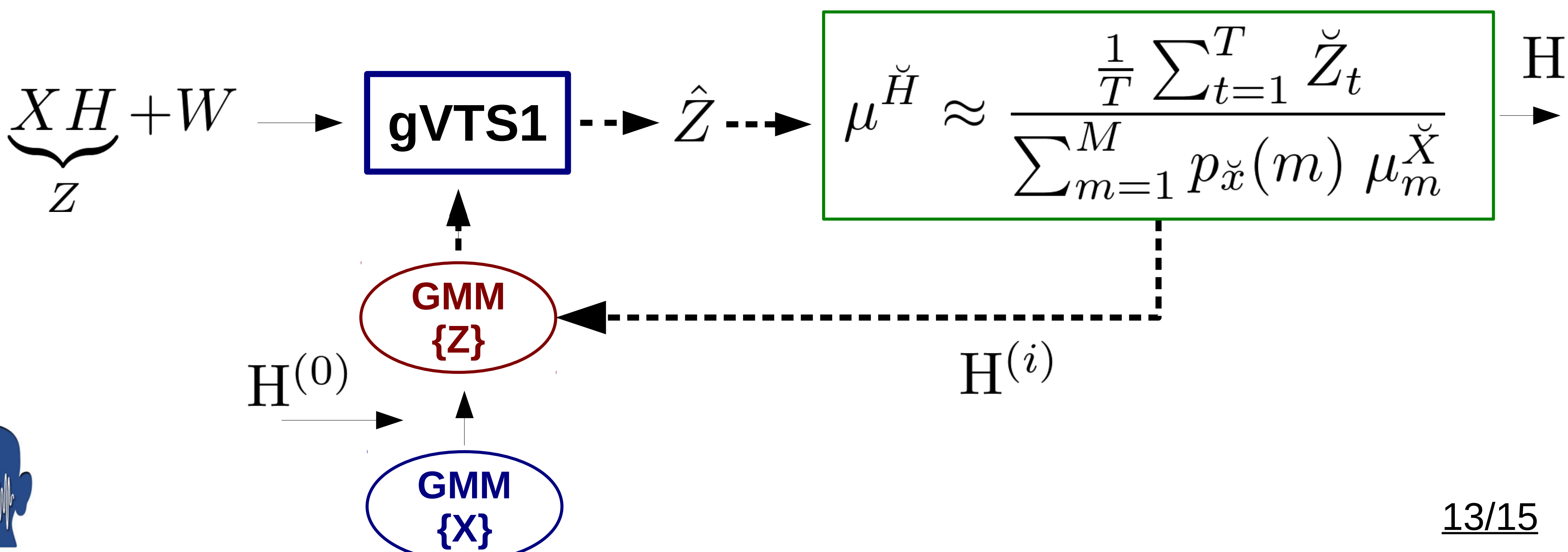
Additive Noise Effect

$$\mathbb{E}\{\frac{\breve{Y}}{\breve{X}}\} = \mathbb{E}\{(\frac{XH+W}{X})^{\alpha}\} \approx \mu^{\breve{H}} + \mathbb{E}\{\frac{\breve{W}}{\breve{X}}\}$$



Additive Noise Effect

$$\mathbb{E}\{\frac{\breve{Y}}{\breve{X}}\} = \mathbb{E}\{(\frac{XH+W}{X})^{\alpha}\} \approx \mu^{\breve{H}} + \mathbb{E}\{\frac{\breve{W}}{\breve{X}}\}$$



Channel Estimation Pseudocode

0. Initialise **H**

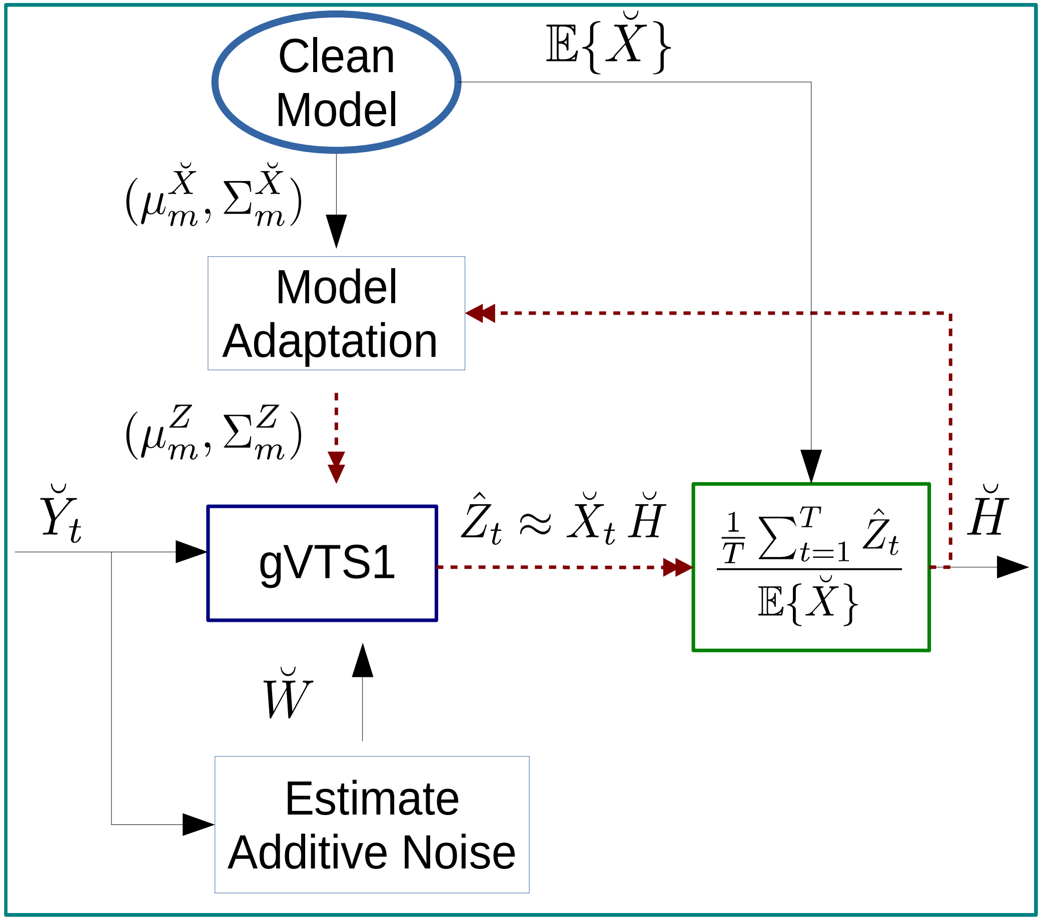
1. Adapt Clean Model with **H**

2. gVTS for Additive Noise

3. Update **H**

4. If not converged **GO TO 1**

RETURN H



Experimental Results
Aurora4 -- WER

Feature	A	B	C	D	Ave_1	Ave_2
MFCC-Clean	6.8	33.4	23.8	50.2	38.0	28.6
MFCC-Multi1	9.1	18.4	23.4	35.9	25.6	21.7
MFCC-Multi2	10.7	17.0	19.1	31.3	22.8	19.5
VTS1-FBE	6.4	21.9	22.2	39.2	28.2	22.4
gVTS1-0.05	6.3	19.9	20.6	36.9	26.2	20.9
gVTS2-0.05-3	6.5	20.3	**14.4**	34.2	24.9	**18.9**

A: Clean
B: Additive
C: Channel
D: Additive+Channel

$$Ave_1 = \frac{A + 6B + C + 6D}{14}$$

$$Ave_2 = \frac{A + B + C + D}{4}$$

Experimental Results

Aurora4 -- WER

Feature	A	B	C	D	Ave_1	Ave_2
MFCC-Clean	6.8	33.4	23.8	50.2	38.0	28.6
MFCC-Multi1	9.1	18.4	23.4	35.9	25.6	21.7
MFCC-Multi2	10.7	17.0	19.1	31.3	22.8	19.5
VTS1-FBE	6.4	21.9	22.2	39.2	28.2	22.4
gVTS1-0.05	6.3	19.9	20.6	36.9	26.2	20.9
gVTS2-0.05-3	6.5	20.3	**14.4**	34.2	24.9	**18.9**

A: Clean
B: Additive
C: Channel
D: Additive+Channel

Abs: 4.7%
Rel: 24.6%

Abs: 0.6%
Rel: 3.1%

WER Reduction

Wrap-up

– This Talk was about ...
  - Vector Taylor Series for Robust ASR
  - Generalised VTS
  - Channel Noise Estimation

– Future Works:
  - Extension to the Phase/Group Delay domains
  - Further optimisation of the channel estimation

SPandH

That's it!

- Thanks for your attention
- Q&A

Appendices

1. (g)VTS Pseudocode
2. Effect of GenLog on WER
    2.1. Aurora-2 → Clean (0-20 dB)
    2.2. Aurora-4 → Clean, Multi1, Multi2
3. Why Non-linear Transform?
4. Channel Estimation
    4.1. Initialisation and iteration effect
    4.2. ALL
5. Phase Factor
6. MMSE vs MAP

(g)VTS Pseudocode

0. GMM of ***CLEAN***

– *For each utterance ...*

1. Apply the compression function (Log of GenLog)

2. Factor out the ***CLEAN*** and compute the distortion function

3. Estimate the ***Noise***
  3.1. Additive
  3.2. Channel

4. Linearise using **VTS**
  4.1. Points → means of Gaussians
  4.2. Jacobians → partial derivatives

5. Estimate *CLEAN* using **MMSE** ← 3 assumptions

App. 1

Effect of GenLog on WER Aurora-2

Feature	α	TestSet A	TestSet B	TestSet C
MFCC	$log \leftrightarrow 0$	66.2	71.4	64.9
gMFCC	0.01	68.0	72.2	69.7
gMFCC	0.05	74.5	76.7	76.0
gMFCC	0.075	**75.4**	**76.2**	**76.9**
gMFCC	0.1	73.3	74.3	74.5
gMFCC	0.15	70.0	71.4	68.8
gMFCC	0.2	67.2	69.3	63.2

Aurora-2 (Accuracy, Average 0-20 dB)

Effect of GenLog on WER Aurora-4

Feature	α	A	B	C	D	Ave_1	Ave_2
MFCC-Clean	log$\leftrightarrow$ 0	6.8	33.7	23.6	49.9	38.0	28.6
gMFCC	0.05	6.9	25.5	23.7	43.1	31.6	24.8
gMFCC	0.075	7.7	22.9	24.3	40.7	29.6	23.9
gMFCC	0.1	7.8	22.2	25.7	40.4	29.2	24.0
MFCC-Multi1	log	9.1	18.4	23.4	35.9	25.6	21.7
gMFCC	0.05	9.3	16.6	23.9	34.5	24.2	21.1
gMFCC	0.075	9.6	16.1	25.4	34.4	24.1	21.4
gMFCC	0.1	10.2	16.0	26.0	34.8	24.3	21.7
MFCC-Multi2	log	10.7	17.0	19.1	31.3	22.8	19.5
gMFCC	0.05	11.0	16.1	19.9	30.3	22.1	19.4
gMFCC	0.075	11.2	16.3	20.5	30.6	22.4	19.7
gMFCC	0.1	11.3	16.7	21.6	30.9	22.7	20.1

A: Clean
B: Additive
C: Channel
D: Additive+Channel

$$Ave_1 = \frac{A + 6B + C + 6D}{14}$$

$$Ave_2 = \frac{A + B + C + D}{4}$$

Taking Log after applying GenLog ?

$$\tilde{Y} = \tilde{X} + \underbrace{\tilde{H} + log\{1 + e^{\tilde{W} - \tilde{X} - \tilde{H}}\}}_{\tilde{G}(\tilde{X},\tilde{W},\tilde{H})}$$

$$\tilde{Z} = logZ$$

$$\breve{Y} \; = \; \breve{X} \; \underbrace{\breve{H} \; (1 + (\frac{\breve{W}}{\breve{X}\breve{H}})^{\frac{1}{\alpha}})^{\alpha}}_{\breve{G}(\breve{X},\breve{W},\breve{H})}$$

$$\breve{Z} = Z^{\alpha}$$

$$\dot{Y} = \dot{X} + \underbrace{\dot{H} + log\{1 + e^{\dot{W} - \dot{X} - \dot{H}}\}}_{\dot{G}(\dot{X},\dot{W},\dot{H})}$$

$$\dot{Z} = \alpha \; logZ$$

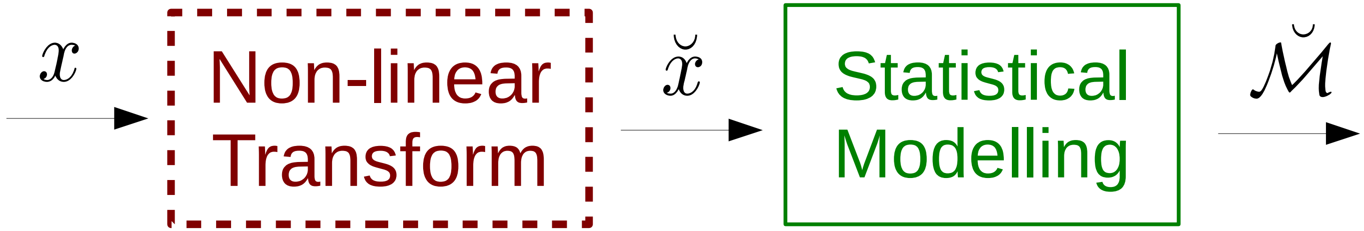
App. 3

Why Non-linear Transform?

$$\hat{Z} = logZ$$

$$\breve{Z} = Z^{\alpha}$$

$$\dot{Z} = \alpha \, logZ$$



Why Non-linear Transform?

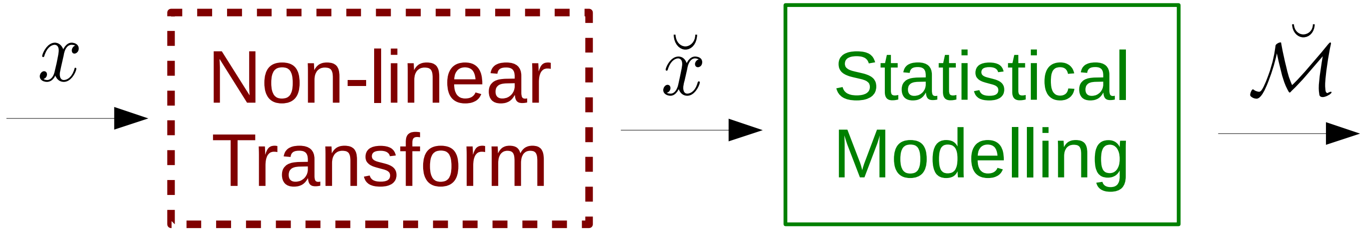
Non-Linear

$$\hat{Z} = log Z$$

$$\breve{Z} = Z^{\alpha}$$

Linear

$$\dot{Z} = \alpha \, log Z$$



Why Non-linear Transform?

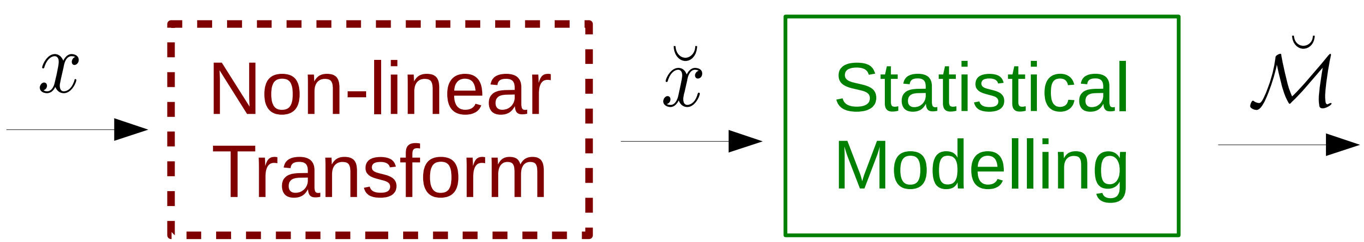
$$\hat{Z} = log Z$$

Non-Linear

$$\breve{Z} = Z^{\alpha}$$

Linear

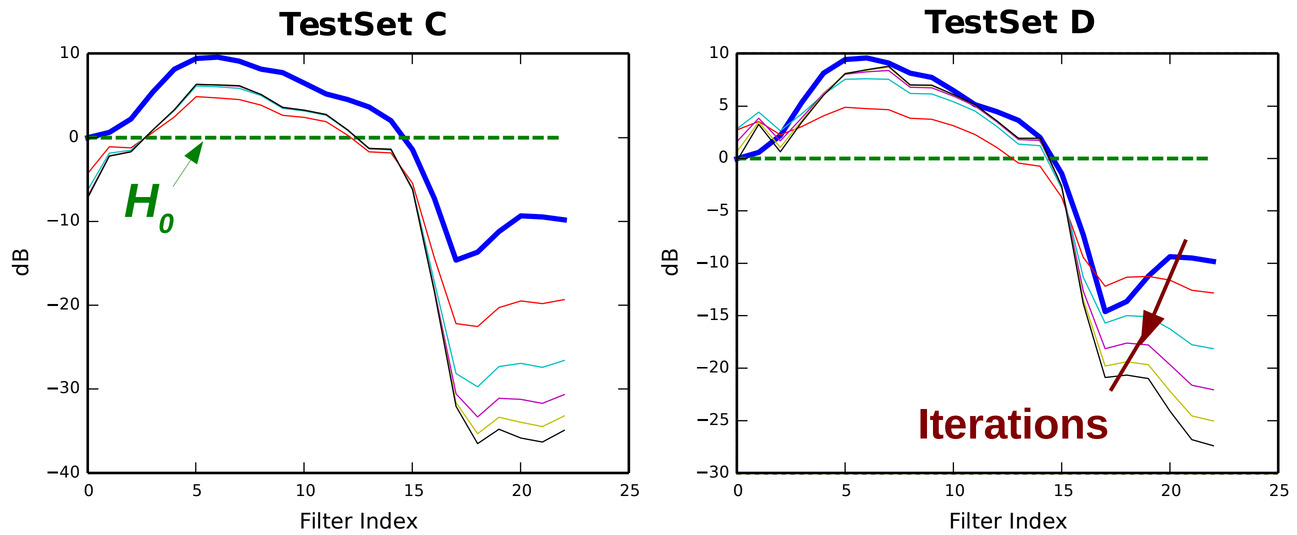
$$\dot{Z} = \alpha \, log Z$$



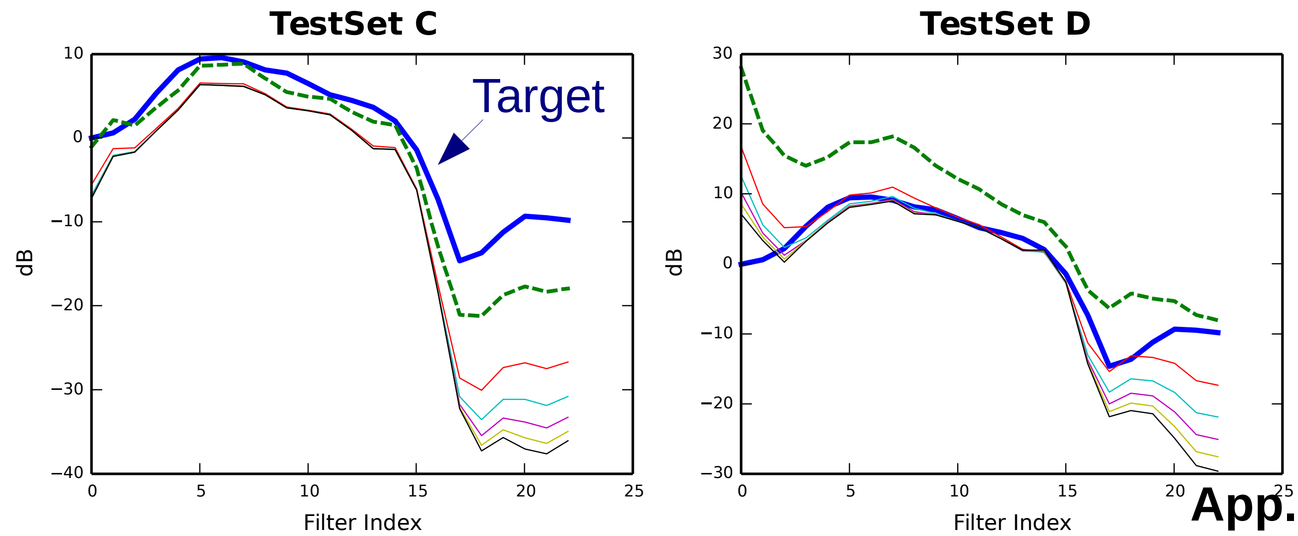
Linear transform does not change the family a RV belongs to !

Channel Estimation
Initialisation and Iteration effect

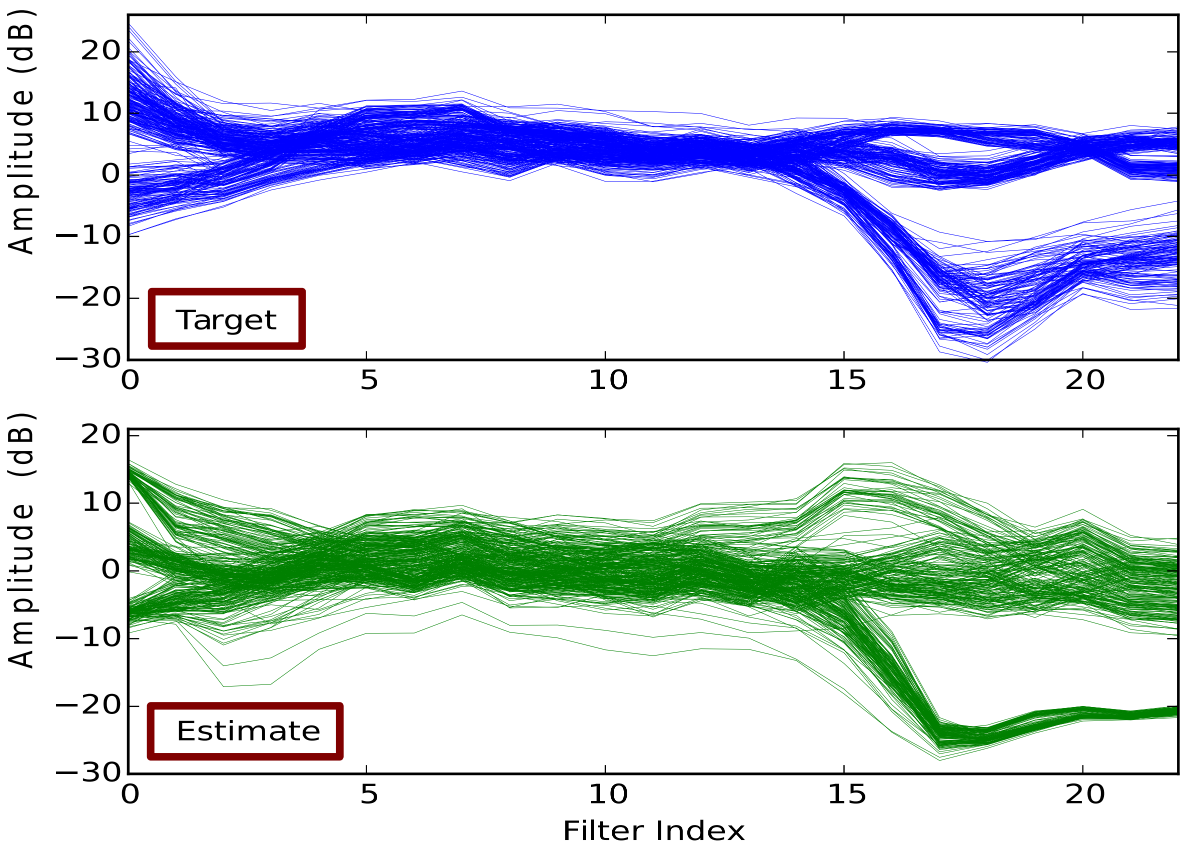
$$\breve{H}_0 = 1$$



$$\breve{H}_0 = \frac{\breve{Y}}{\breve{X}}$$



Channel Estimation -- ALL



$$\mu^{\breve{H}} = \mathbb{E}\{\breve{H}\} = \mathbb{E}\{\frac{\breve{Y}_u}{\breve{X}_u}\}$$

Y = XH ← Test Set C
X ← Test Set A

Phase Factor

Periodogram

$$Y_k \;=\; X_k H_k \;+\; W_k \;+\; 2\sqrt{X_k H_k W_k} cos(\theta_{X_k} + \theta_{H_k} - \theta_{W_k})$$

$$\lambda_k = cos(\theta_{X_k} + \theta_{H_k} - \theta_{W_k}) = \frac{Y_k \;-\; X_k H_k \;-\; W_k}{2\sqrt{X_k H_k W_k}}$$

FBE, l^{th} filter

$$\lambda_l = \frac{Y_l \;-\; X_l H_l \;-\; W_l}{2\sqrt{X_l H_l W_l}}$$

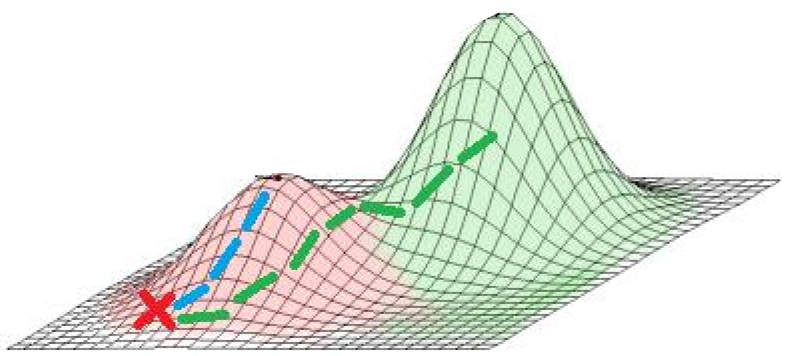
$$\lambda_l \sim \mathcal{N}(0, \sigma_l^2) \Rightarrow \lambda \sim \mathcal{N}(0, \Sigma^\lambda)$$

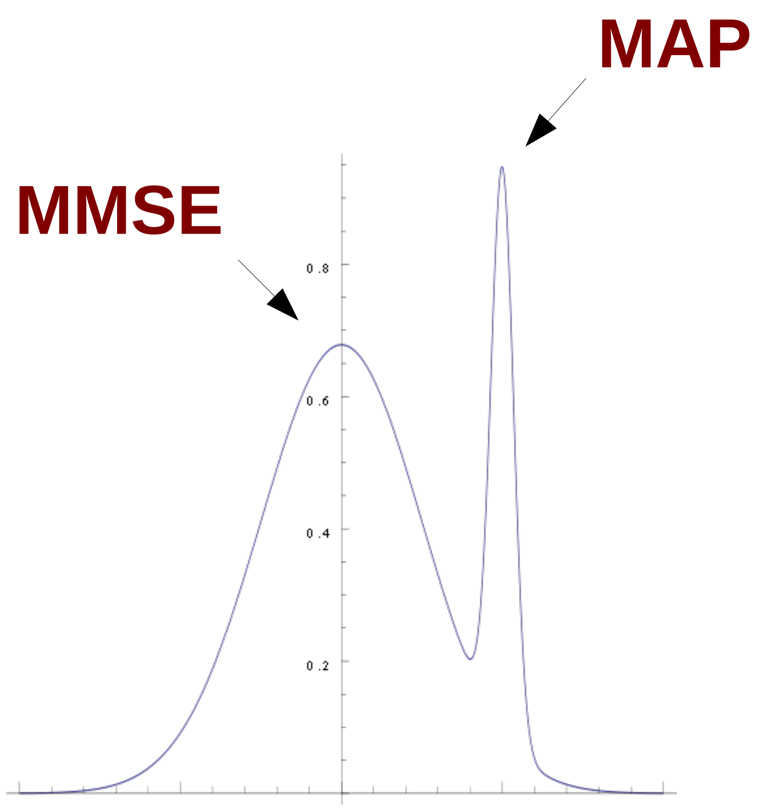
$$J^\lambda_{gVTS} = \frac{\partial \breve{Y}}{\partial \lambda} = \frac{\partial \breve{Y}}{\partial Y}\frac{\partial Y}{\partial \lambda} = 2\alpha \frac{\mu_m^{\breve{X}} \mu^{\breve{H}} \sqrt{V_m}}{(1 + V_m)^{1-\alpha}}$$

MMSE vs MAP

$$\hat{x}_{MMSE} = \mathbb{E}[x|y] = \int x \, P(x|y) \, dx$$

$$\hat{x}_{MAP} = \arg\max_{x} P(x|y)$$





MAP requires
GLOBAL maximum!

App. 6