Speech Enhancement in the DFT Domain Using Supergaussian Priors

Rainer Martin

Technical University of Braunschweig

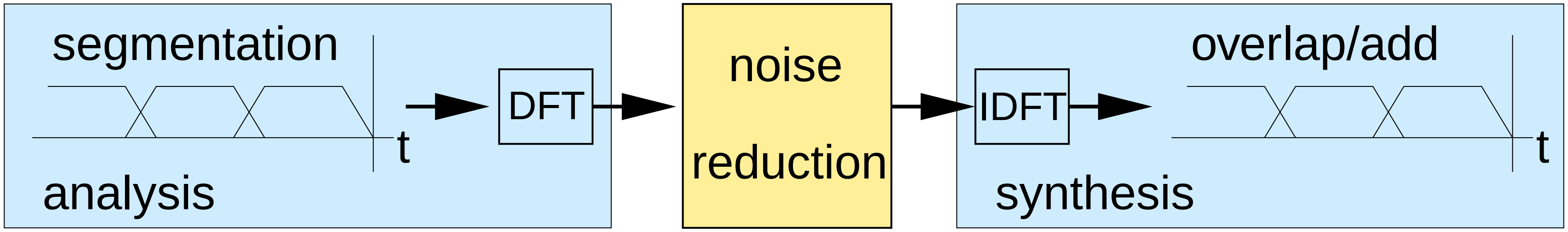
July 2003

Outline

- **Noise Reduction in the Spectral Domain**
- **Statistical Models and Estimation**
- **Experimental Results**
- **Conclusions**

Noise Reduction in the Spectral Domain

▶ Spectral analysis – noise reduction – synthesis:



▶ Advantages of spectral processing:

- good separation of speech and noise
- decorrelation of spectral components
- integration of psychoacoustic models

MMSE Estimation

- **Optimal estimate for independent real and imaginary parts:**

$$E\{S \mid Y\} = E\{S_R \mid Y_R\} + jE\{S_I \mid Y_I\}$$

- **Estimation of either the real or the imaginary part:**

$$E\{S_\diamond \mid Y_\diamond\} = \int_{-\infty}^{\infty} S_\diamond p(S_\diamond \mid Y_\diamond) dS_\diamond$$

- **Application of Bayes theorem:**

$$E\{S_\diamond \mid Y_\diamond\} = \frac{1}{p(Y_\diamond)} \int_{-\infty}^{\infty} S_\diamond p(Y_\diamond \mid S_\diamond) p(S_\diamond) dS_\diamond$$

- **What is the appropriate prior density $p(S_\diamond)$?**

Prior Densities for Real and Imaginary Part

- **Gaussian pdf:**

$$p(S_\diamond) = \frac{1}{\sqrt{\pi}\sigma_s} \exp(-\frac{S_\diamond^2}{\sigma_s^2})$$

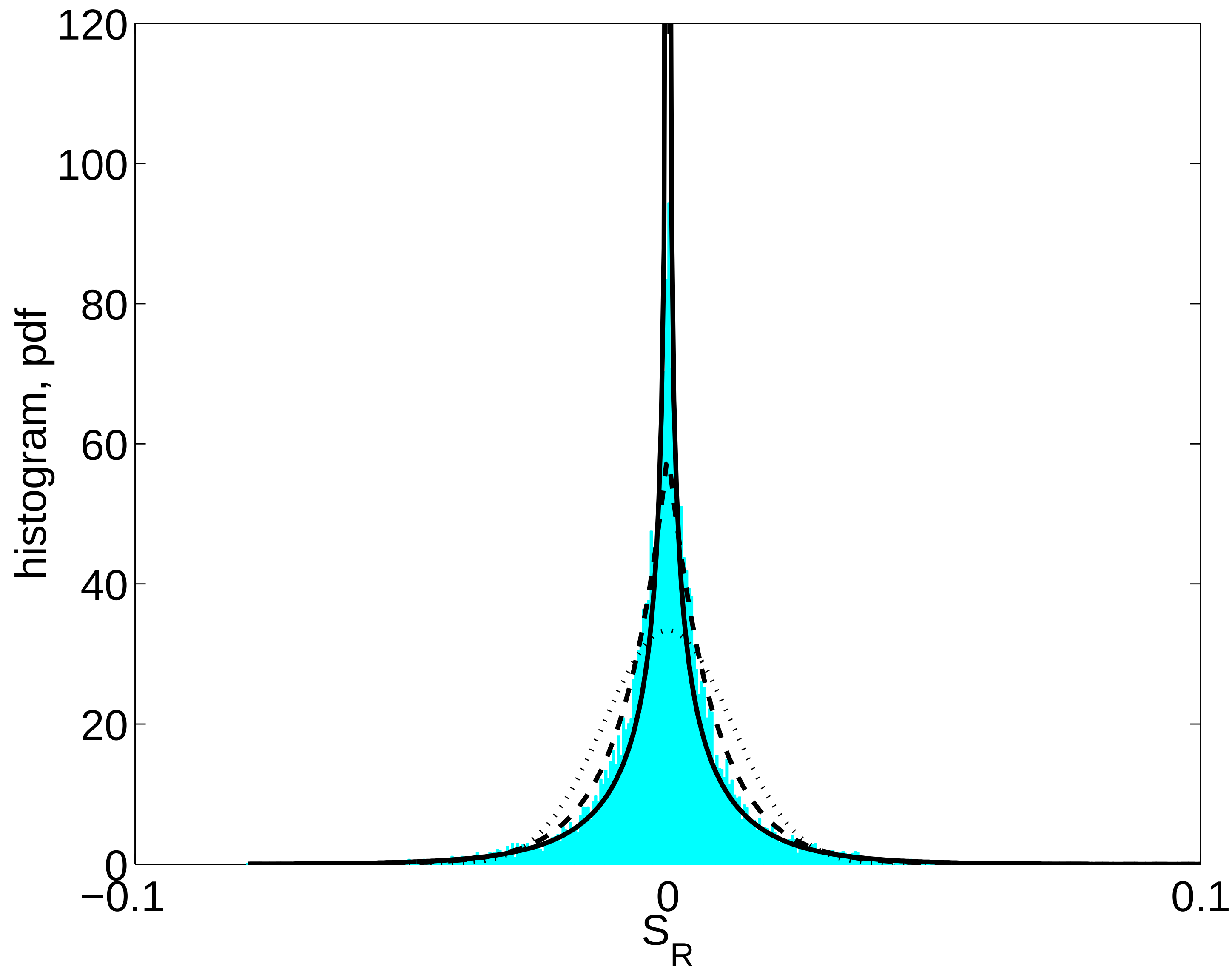
- **Laplacian pdf:**

$$p(S_\diamond) = \frac{1}{\sigma_s} \exp(-\frac{2|S_\diamond|}{\sigma_s})$$

- **Gamma pdf:**

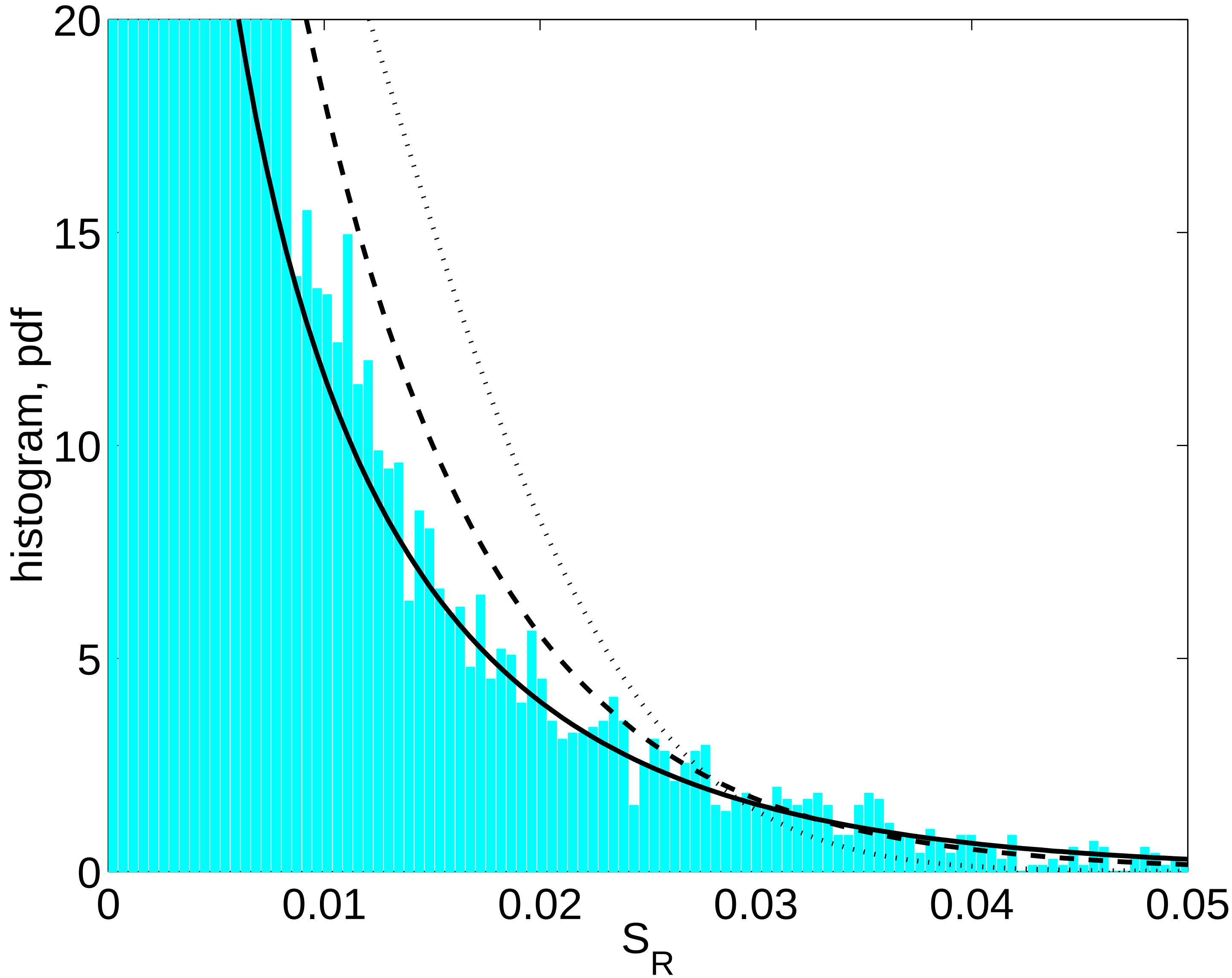
$$p(S_\diamond) = \frac{\sqrt[4]{3}}{2\sqrt{\pi\sigma_s}\sqrt[4]{2}} |S_\diamond|^{-\frac{1}{2}} \exp(-\frac{\sqrt{3}|S_\diamond|}{\sqrt{2}\sigma_s})$$

Histogram of DFT Coefficients for Speech



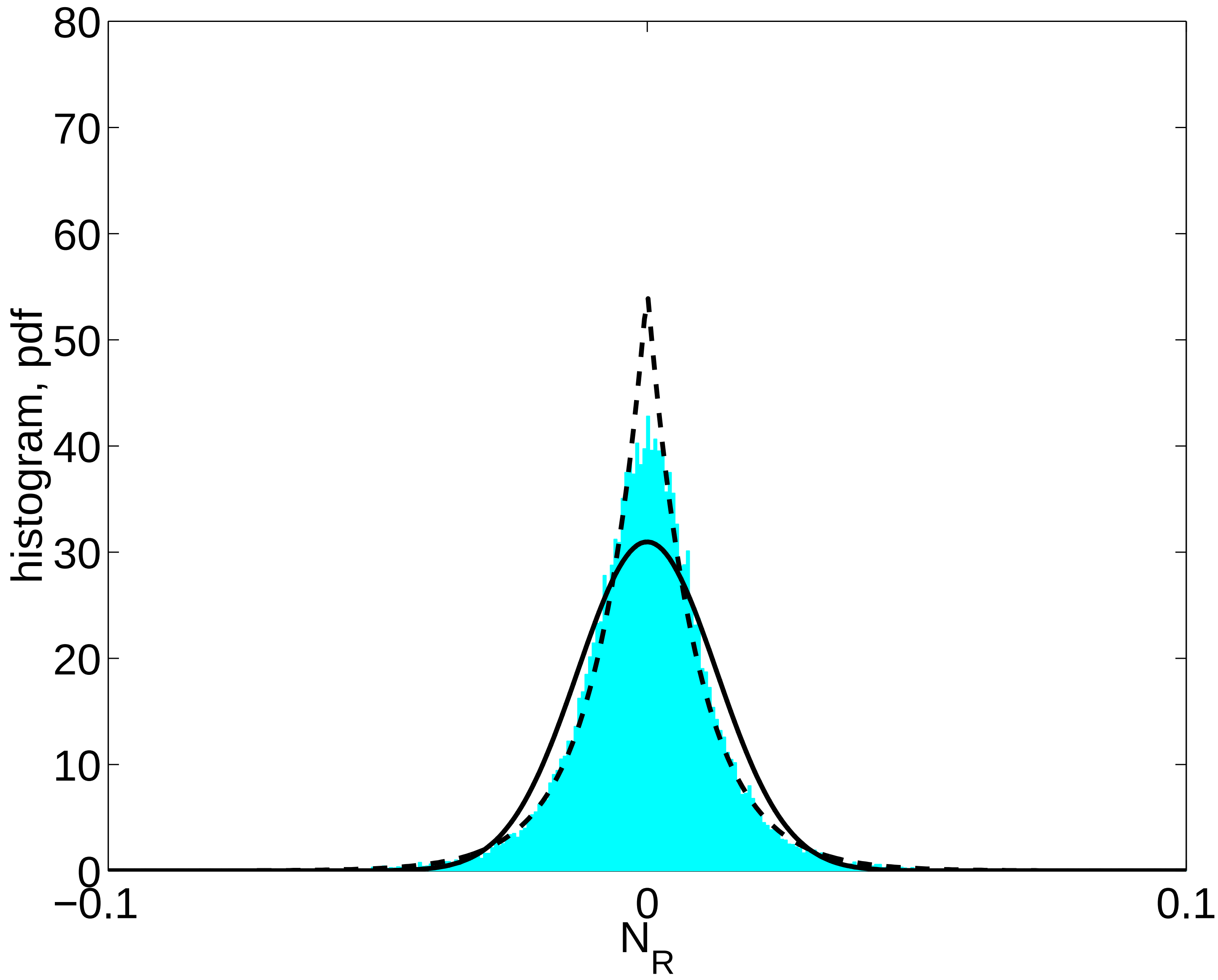
dotted: Gaussian pdf dashed: Laplacian pdf solid: Gamma pdf

Histogram of Speech Coefficients (enlarged)



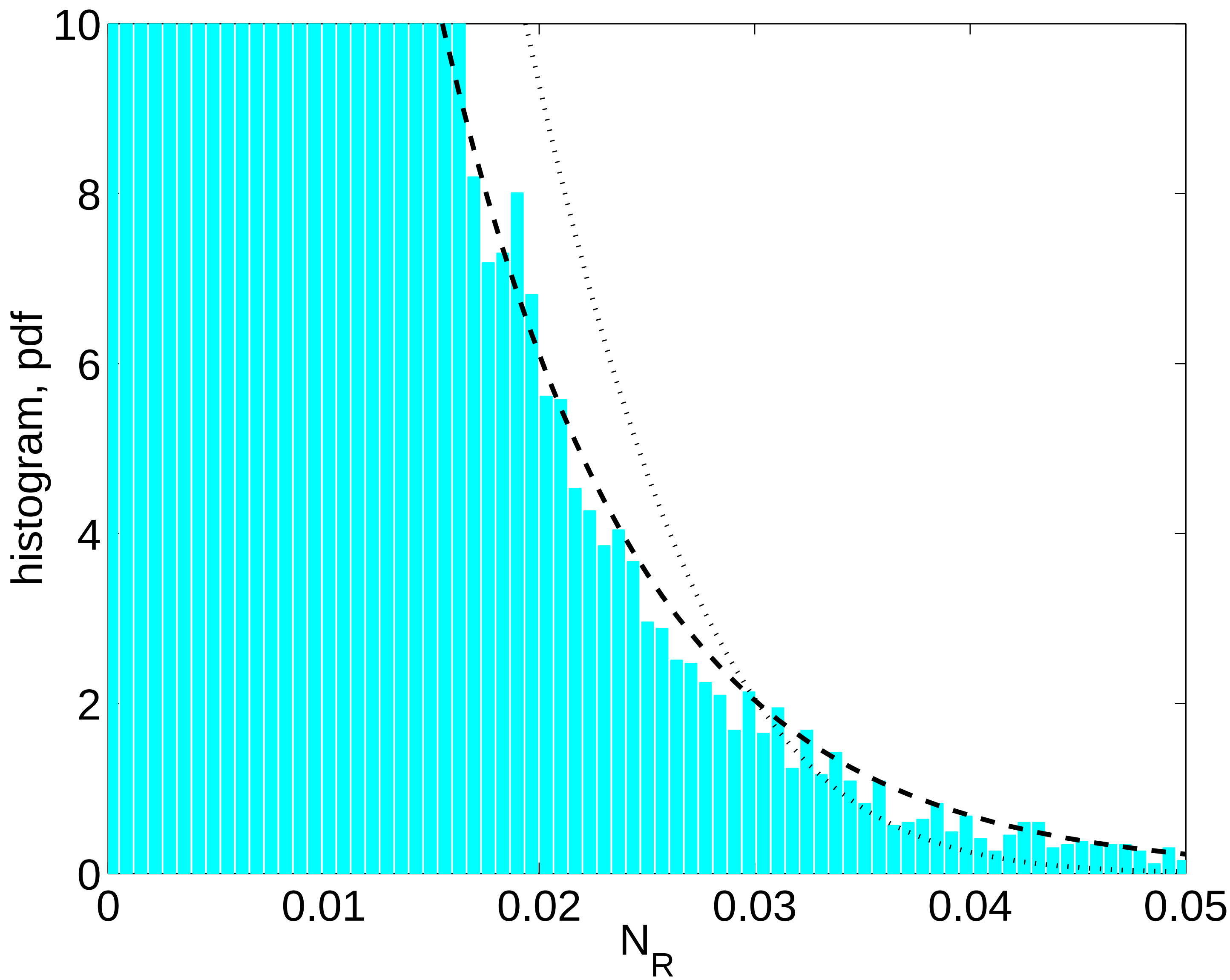
dotted: Gaussian pdf dashed: Laplacian pdf solid: Gamma pdf

Histogram of DFT Coefficients for Car Noise



dotted: Gaussian pdf   dashed: Laplacian pdf

Histogram of Car Coefficients (enlarged)



dotted: Gaussian pdf dashed: Laplacian pdf

MMSE Estimator for Gamma Speech Priors

$$E\{S \mid Y\} = E\{S_R \mid Y_R\} + jE\{S_I \mid Y_I\}$$

$$E\{S_R \mid Y_R\}$$

$$= \frac{\sigma_n \big(\exp(G_{R-}^2/2) D_{-1.5}(\sqrt{2} G_{R-}) - \exp(G_{R+}^2/2) D_{-1.5}(\sqrt{2} G_{R+}) \big)}{2\sqrt{2} \Big(\exp(G_{R-}^2/2) D_{-0.5}(\sqrt{2} G_{R-}) + \exp(G_{R+}^2/2) D_{-0.5}(\sqrt{2} G_{R+}) \Big)}$$

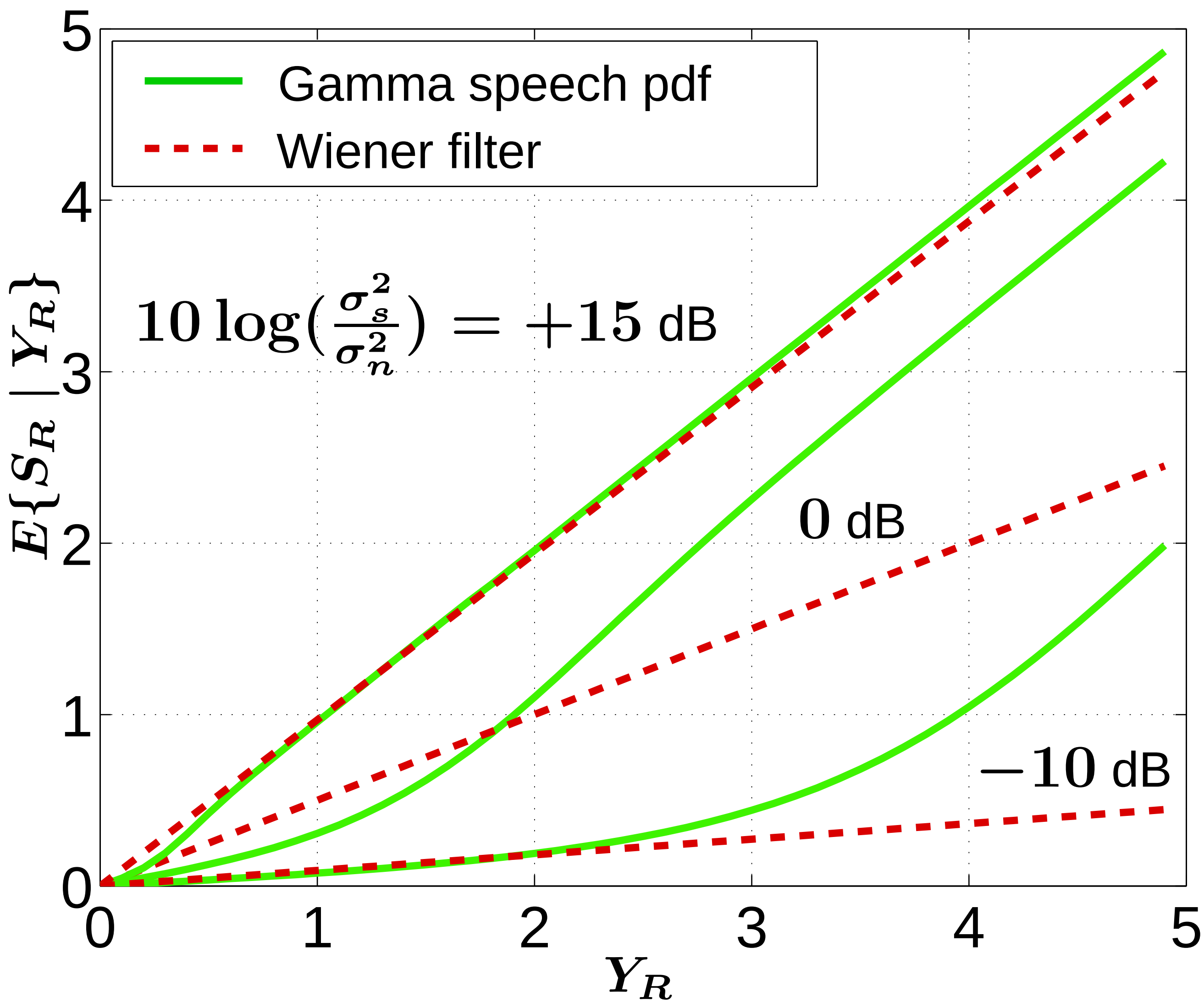
$$G_{R\pm} = \frac{\sqrt{3}\sigma_n}{2\sqrt{2}\sigma_s} \pm \frac{Y_R}{\sigma_n} \qquad G_{I\pm} = \frac{\sqrt{3}\sigma_n}{2\sqrt{2}\sigma_s} \pm \frac{Y_I}{\sigma_n}$$

$D_{-0.5}(z)$ and $D_{-1.5}(z)$ are parabolic cylinder functions.

$$E\{S_I \mid Y_I\}$$

$$= \frac{\sigma_n \big(\exp(G_{I-}^2/2) D_{-1.5}(\sqrt{2} G_{I-}) - \exp(G_{I+}^2/2) D_{-1.5}(\sqrt{2} G_{I+}) \big)}{2\sqrt{2} \Big(\exp(G_{I-}^2/2) D_{-0.5}(\sqrt{2} G_{I-}) + \exp(G_{I+}^2/2) D_{-0.5}(\sqrt{2} G_{I+}) \Big)}$$

Gaussian Noise and Gamma Speech Prior



$$\sigma_s^2 + \sigma_n^2 = 2$$

Gaussian Noise and Laplacian Speech Prior

$$E\{S_R \mid Y_R\}$$

$$= \frac{\sigma_n \left[L_{R+} \exp(L_{R+}^2) \operatorname{erfc}(L_{R+}) - L_{R-} \exp(L_{R-}^2) \operatorname{erfc}(L_{R-}) \right]}{\exp(L_{R+}^2) \operatorname{erfc}(L_{R+}) + \exp(L_{R-}^2) \operatorname{erfc}(L_{R-})}$$

and

$$E\{S_I \mid Y_I\}$$

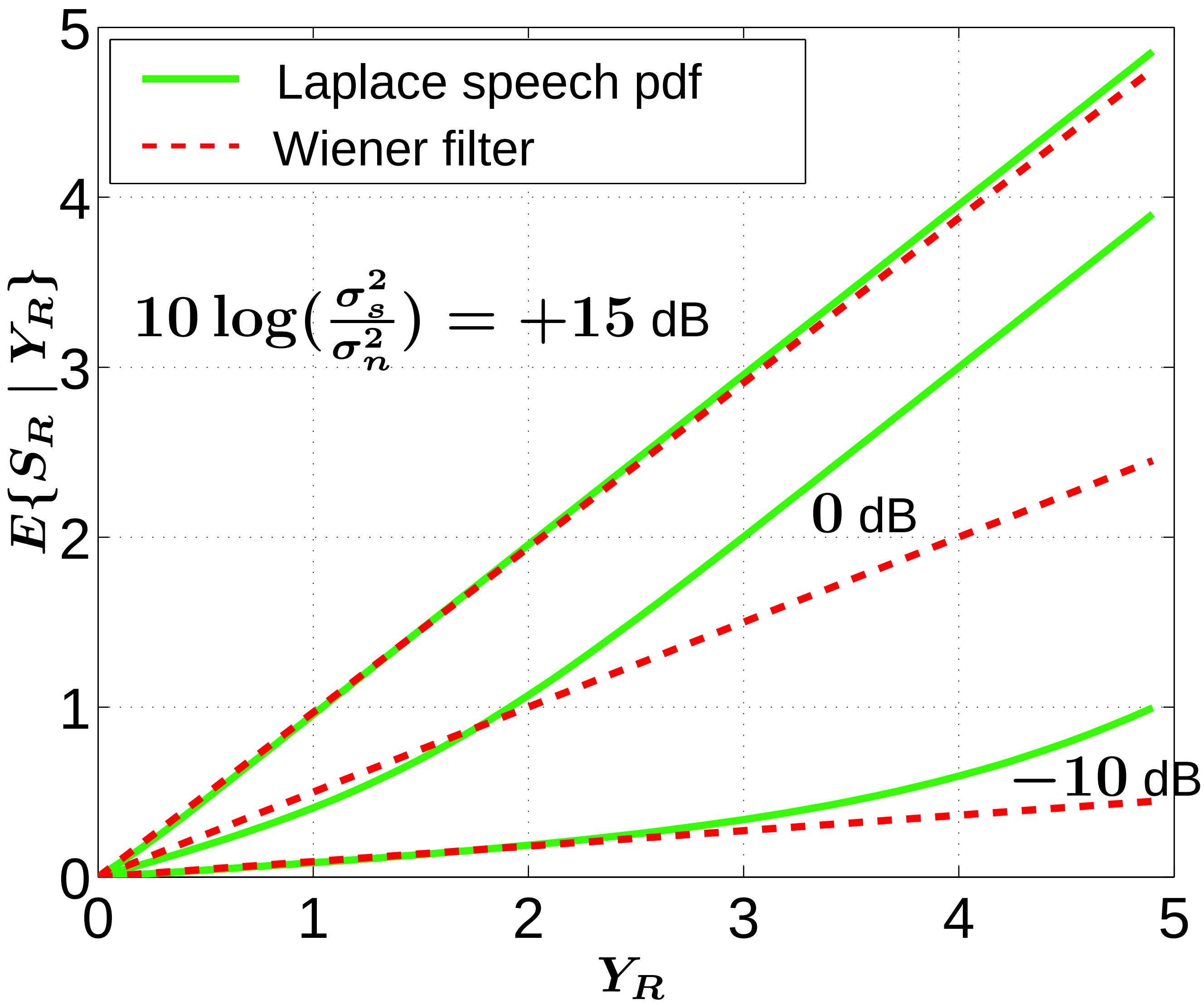
$$= \frac{\sigma_n \left[L_{I+} \exp(L_{I+}^2) \operatorname{erfc}(L_{I+}) - L_{I-} \exp(L_{I-}^2) \operatorname{erfc}(L_{I-}) \right]}{\exp(L_{I+}^2) \operatorname{erfc}(L_{I+}) + \exp(L_{I-}^2) \operatorname{erfc}(L_{I-})}$$

with

$$L_{R+} = \frac{\sigma_n}{\sigma_s} + \frac{Y_R}{\sigma_n} \qquad L_{R-} = \frac{\sigma_n}{\sigma_s} - \frac{Y_R}{\sigma_n}$$

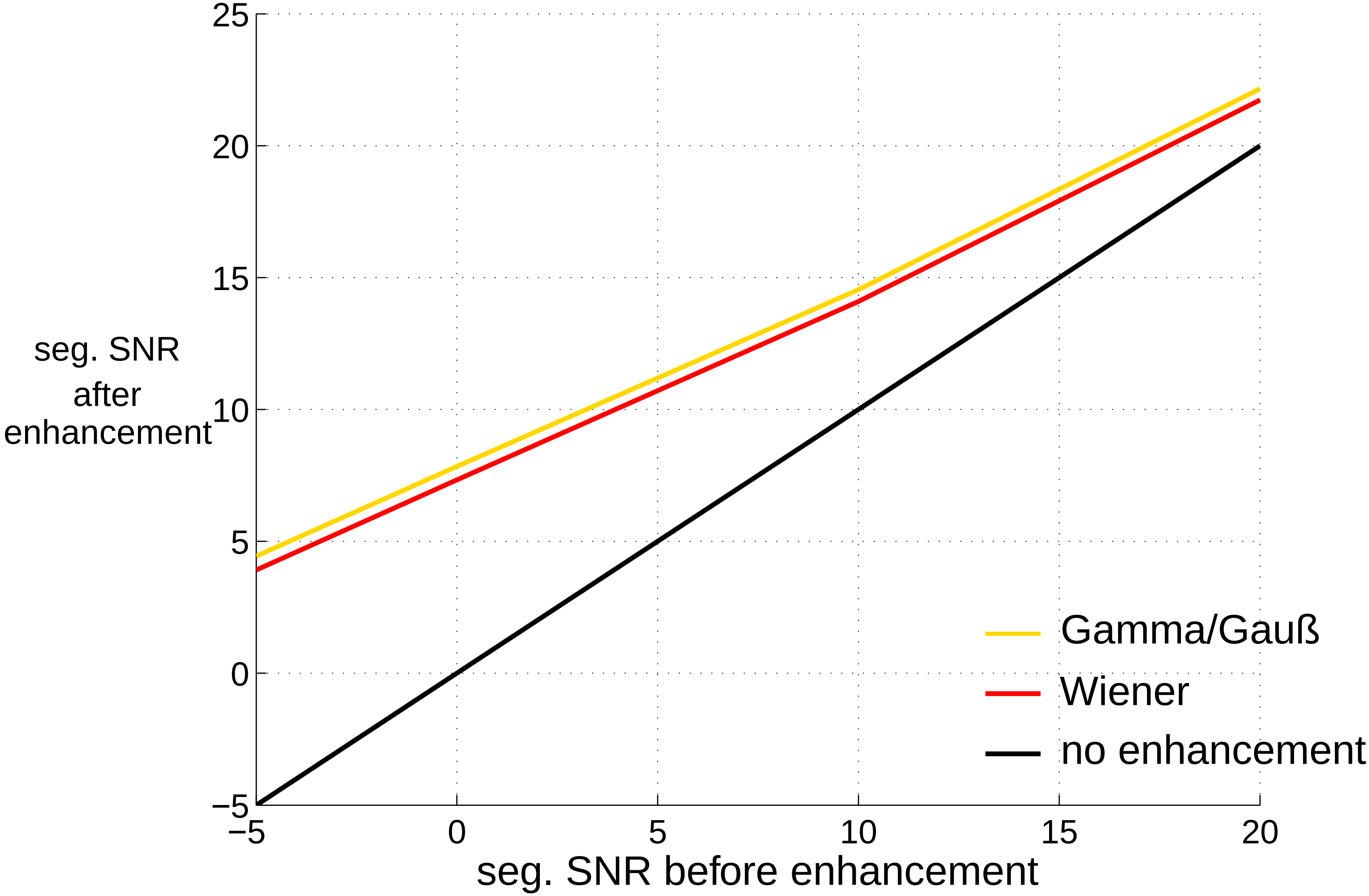
$$L_{I+} = \frac{\sigma_n}{\sigma_s} + \frac{Y_I}{\sigma_n} \qquad L_{I-} = \frac{\sigma_n}{\sigma_s} - \frac{Y_I}{\sigma_n}$$

Gaussian Noise and Laplacian Speech Prior

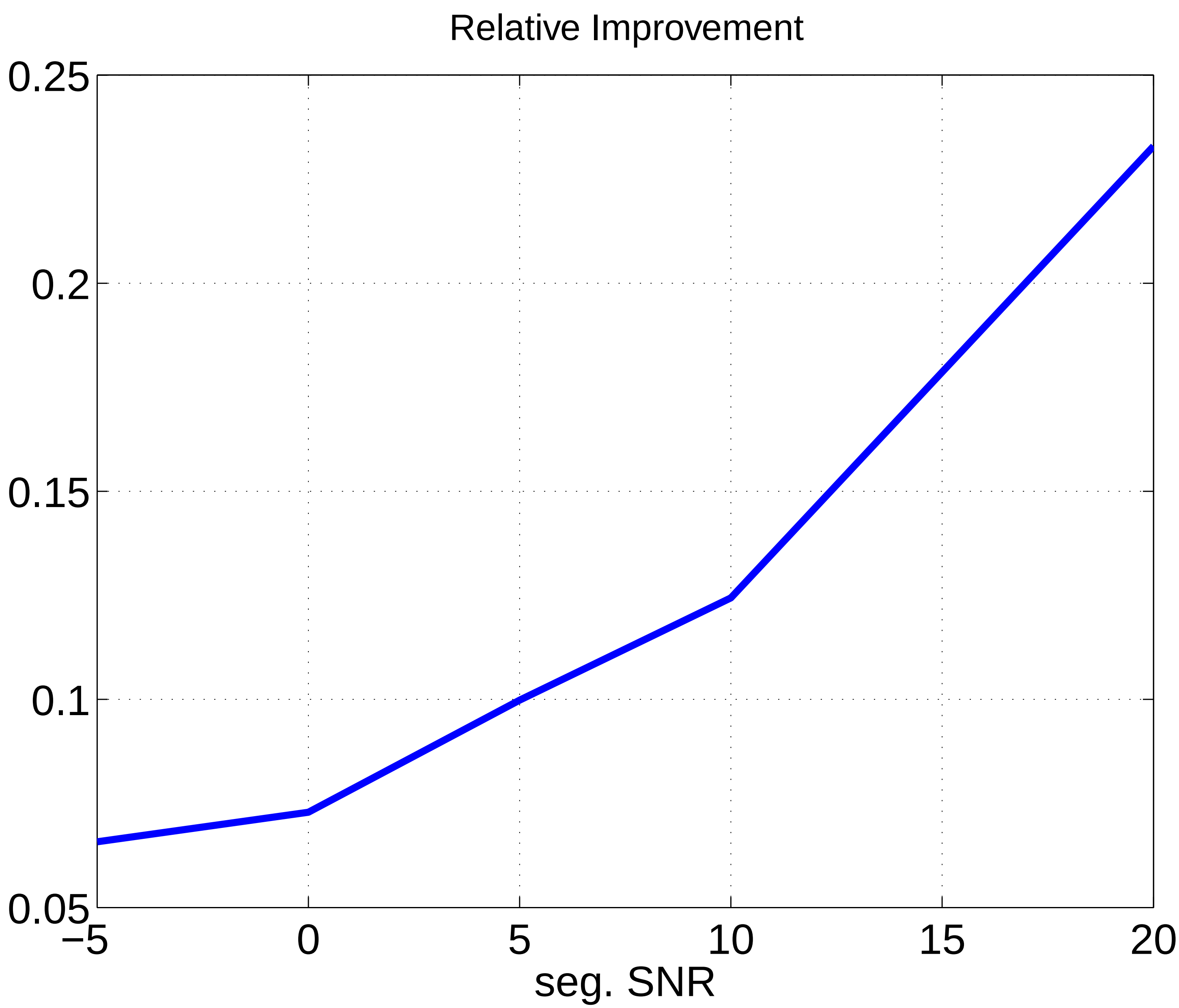


$$\sigma_s^2 + \sigma_n^2 = 2$$

Segmental SNR Improvement (White Noise)



Relative Improvement w.r.t. Wiener Filter



Conclusions

- **Analytic solution for MMSE estimation of DFT coefficients with Gamma or Laplacian speech priors.**
  - Segmental SNR improvement of 0.3 ... 0.5 dB with respect to Wiener Filter;
  - Relative improvement of segmental SNR of 5 - 25 %;
  - New estimator delivers audibly more transparent speech;