Digital Audio and Speech Processing (EE4182)

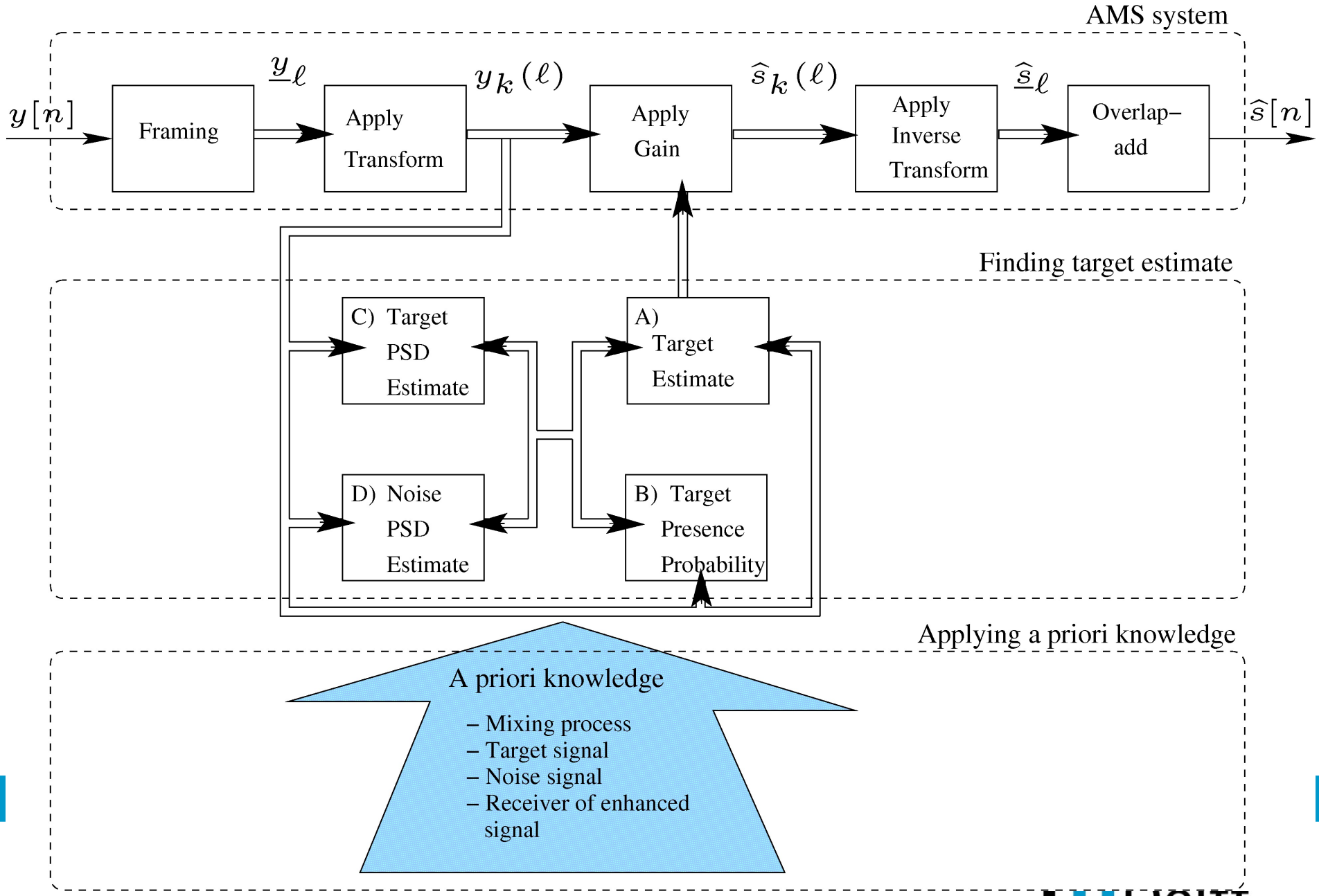
Richard C. Hendriks

20/4/2021

TU Delft

Delft University of Technology

Overview of single-channel NR algorithm



Previous Lecture:

What kind of processing?

For example power spectral subtraction

$$\widehat{|s'_k(l)|} = \left(\max \left\{ 1 - \frac{E\left[|N_k(l)|^2\right]}{\overline{|y_k(l)|^2}}, 0.2 \right\} \right)^{\frac{1}{2}} |y_k(l)|.$$

with $\overline{|y_k(l)|^2} = \frac{1}{L} \sum_{m=l-L+1}^{l} |y_k(l)|^2$.

Power spectral subtraction is very simple, but

- It is rather heuristic (although it can be shown to follow as a ML estimate).
- It does not optimize a specific distortion measure.
- Quality of the noise reduced speech is not great.
- Noise reduced speech contains a lot of musical noise.

Today:

- More advanced estimators that are
  - Mathematically more solid
  - Derived under distributional assumptions that match distribution of speech DFT coefficients.
- An alternative way to the Bartlett estimate for speech PSD estimation that reduces musical noise and increases quality.

Wiener Smoother - Time Domain

Model: $Y(n) = S(n) + N(n)$

How to compute $\hat{S}(n)$?

Let us assume that we compute $\hat{S}(n)$ as $\hat{S}(n) = \sum_{k=0}^{N} h(k)Y(n-k)$.

with

$$h(k) = \begin{cases} \text{something} & 0 \leq k < N \\ 0 & \text{otherwise} \end{cases}$$

How to find the optimal impulse response $h(k)$?

Wiener Smoother - Time Domain

Compute the mean-square error (MSE) optimal filter coefficients $h(k)$:

$$\min_{h(m)} E\left[\left(S(n)-\sum_{k=0}^{N} h(k)Y(n-k)\right)^2\right] \quad \forall \ m$$

$$\frac{d}{dh(m)} E\left[\left(S(n)-\sum_{k=0}^{N} h(k)Y(n-k)\right)^2\right] =$$

$$-2E\left[\left(S(n)-\sum_{k=0}^{N} h(k)Y(n-k)\right)Y(n-m)\right] = 0$$

$$E\left[S(n)Y(n-m)\right] - E\left[\sum_{k=0}^{N} h(k)Y(n-k)Y(n-m)\right] = 0$$

TU Delft

Wiener Smoother - Time Domain

$$E\left[S(n)Y(n-m)\right] - E\left[\sum_{k=0}^{N} h(k)Y(n-k)Y(n-m)\right] = 0$$

$$E\left[S(n)Y(n-m)\right] = E\left[\sum_{k=0}^{N} h(k)Y(n-k)Y(n-m)\right]$$

$$E\left[S(n)Y(n-m)\right] = \sum_{k=0}^{N} h(k)E\left[Y(n-k)Y(n-m)\right]$$

Cross-correlation

auto-correlation

$$R_{SY}(m) = \sum_{k=0}^{N} h(k)R_{YY}(m-k)$$

TU Delft

Wiener Smoother - Time Domain

$$R_{SY}(m) = \sum_{k=0}^{N} h(k) R_{YY}(m-k)$$

Write this in Matrix form:

$$\underbrace{\begin{pmatrix} R_{SY}(0) \\ R_{SY}(1) \\ \vdots \\ R_{SY}(N) \end{pmatrix}}_{\mathbf{r}_{SY}} = \underbrace{\begin{pmatrix} R_{YY}(0) & R_{YY}(-1) & \cdots & R_{YY}(-N) \\ R_{YY}(1) & R_{YY}(0) & \cdots & R_{YY}(-N+1) \\ \vdots & \vdots & \ddots & \vdots \\ R_{YY}(N) & R_{YY}(N-1) & \cdots & R_{YY}(0) \end{pmatrix}}_{\mathbf{R_{YY}}} \underbrace{\begin{pmatrix} h(0) \\ h(1) \\ \vdots \\ h(N) \end{pmatrix}}_{\mathbf{h}}$$

Compact notation: $\mathbf{h} = \mathbf{R}_{YY}^{-1}\mathbf{r}_{SY} \rightarrow \hat{s} = \mathbf{h}^T\mathbf{y}$

TU Delft

The Wiener Smoother - Freq. Domain

Convolution between $h(m)$ and $R_{YY}(m)$!!

$$R_{SY}(j) = \sum_{m=0}^{N} h(m) R_{YY}(j-m)$$

Taking Fourier Transform (and assume long frames).

$$P_{SY,k} = H_k P_{YY,k}$$

$$H_k = \frac{P_{SY,k}}{P_{YY,k}},$$

where

- $P_{SY,k}$ is the cross-power spectral density of $S_t(n)$ and $Y_t(n)$
- $P_{YY,k}$ is the power spectral density of $Y_t(n)$.

TU Delft

The Wiener Smoother - Freq. Domain

An alternative derivation goes via Parseval's theorem:

$$\sum_n (S_t(n) - \underbrace{h(n) * Y_t(n)}_{\hat{S}_t(n)})^2 = \frac{1}{K} \sum_{k=0}^{K-1} |S_k - \underbrace{H_k Y_k}_{\hat{S}_k}|^2$$

we can find the filter H_k in freq. domain by solving

$$\frac{\partial}{\partial H_k} E\{|S_k - H_k Y_k|^2\} = 0$$

for H_k.

The Wiener Smoother - Freq. Domain

Solving

$$\frac{\partial}{\partial H_k} E\{|S_k - H_k Y_k|^2\} = 0$$

leads to

$$H_k = \frac{P_{SY,k}}{P_{YY,k}},$$

where

- $P_{SY,k} = \frac{1}{L} E\{S_k Y_k^*\}$ is the cross-power spectral density of $S_t(n)$ and $Y_t(n)$
- $P_{YY,k} = \frac{1}{L} E\{Y_k Y_k^*\}$ is the power spectral density of $Y_t(n)$.

The Wiener Smoother - Freq. Domain

So far we did not assume anything with respect to correlation between $S_t(n)$ and $N_t(n)$. For the problem at hand we assume that $S_t(n)$ and $N_t(n)$ are uncorrelated:

$$\mathbf{R_{YY}} = \mathbf{R_{SS}} + \mathbf{R_{NN}},$$
$$P_{YY,k} = P_{SS,k} + P_{NN,k}.$$

Time-domain:

$$\mathbf{h} = \mathbf{R}_{YY}^{-1}\mathbf{r}_{SY} = \mathbf{R}_{YY}^{-1}\mathbf{r}_{SS}$$

Frequency-domain:

$$H_k \approx P_{SS,k}/P_{YY,k} = P_{SS,k}/(P_{SS,k} + P_{NN,k}).$$

The Wiener Smoother - Freq. Domain

We compute the estimator in the frequency domain as follows:

$$\begin{aligned}\hat{S}_k &= H_k \cdot Y_k \qquad (\forall\, k) \\ &= \frac{P_{SS,k}}{P_{SS,k} + P_{NN,k}} |Y_k| e^{j \angle Y_k}\end{aligned}$$

Since power spectral densities $P_{SS,k}$ and $P_{NN,k}$ are real-valued by definition, we see that the filter *does not change the phase of the input* Y_k. In other words, using the noisy phase is *optimal* in this situation!

The Wiener Smoother - Freq. Domain

Let us write $H_k = \frac{P_{SS,k}}{P_{SS,k}+P_{NN,k}}$ as

$$H_k = \frac{P_{SS,k}/P_{NN,k}}{P_{SS,k}/P_{NN,k}+1} = \frac{SNR_k}{SNR_k+1}.$$

It follows that

$$H_k \to 1 \text{ for } SNR_k \to \infty$$
$$H_k \to 0 \text{ for } SNR_k \to 0$$

We see that the Wiener filter suppresses spectral regions with low SNR while it does not modify spectral regions where the SNR is high!

The Wiener Smoother - Freq. Domain

How to implement the Wiener Smoother $H_k \approx \frac{P_{SS,k}}{P_{SS,k}+P_{NN,k}}$ (or $\mathbf{h} = \mathbf{R}_{\mathbf{YY}}^{-1}\mathbf{R}_{\mathbf{SY}} = \left(\mathbf{R}_{\mathbf{SS}} + \mathbf{R}_{\mathbf{NN}}\right)^{-1}\mathbf{R}_{\mathbf{SS}}$)? (We don't know $P_{SS,k}$ (or $\mathbf{R}_{\mathbf{SS}}$)).

Let us re-write $H(\omega)$ as follows:

$$H_k \approx \frac{P_{SS,k} + P_{NN,k} - P_{NN,k}}{P_{SS,k} + P_{NN,k}} = 1 - \frac{P_{NN,k}}{P_{YY,k}}.$$

The power spectral density $P_{YY,k}$ may be estimated e.g. through a Bartlett estimate. $P_{NN,k}$ may be estimated using tracking methods (next lecture).

MMSE and Conditional Mean

Question:

- Is the Wiener filter (smoother) the best (in mmse sense) estimator we can find?

Answer(s):

- The Wiener filter is the best *linear* estimator.
- However, if we have a priori information on the pdfs of the random variables involved (Y_k, S_k and N_k in our specific case), we can generally do better!

MMSE and Conditional Mean

We now present a general methodology for deriving *globally* optimal MMSE estimators. Later we apply this to the noise suppression problem.

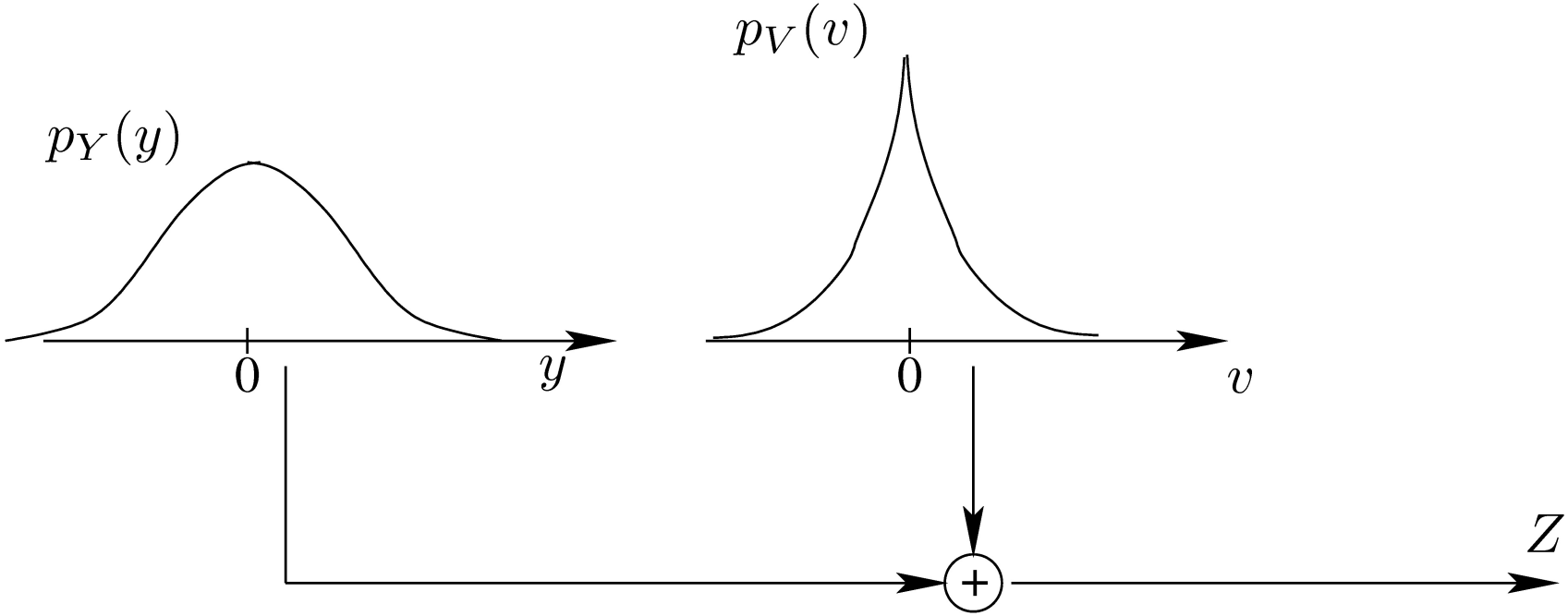
Consider the signal model:

$$Z = Y + V$$

- Y is random variable representing quantity of interest.
- V is random variable representing additive disturbance.
- Z is random variable representing observed quantity.
- Assume (for convenience) $E\{Z\} = E\{Y\} = E\{V\} = 0$.

MMSE and Conditional Mean

Signal Model:



We assume noisy *realization* of Z is generated by

- Drawing realization of Y according to pdf $p_Y(y)$.
- Drawing realization of V according to pdf $p_V(v)$.
- Forming noisy realization $z = y + v$.

MMSE and Conditional Mean

Our goal is to estimate realization of Y in minimum mean-squared error sense:

$$\hat{Y}^* = \arg\min_{\hat{Y}} E\{D(Y,\hat{Y})\} \text{ with } D(Y,\hat{Y}) = (Y-\hat{Y})^2.$$

Obviously, our estimator $\hat{Y}$ is some (unknown) function $g(\cdot)$ of Z, i.e., $\hat{Y} = g(Z)$.

Since $D(Y,\hat{Y})$ is a random variable which is dependent on Z (through $\hat{Y} = g(Z)$) and on Y, we get (by definition)

$$E\{D(Y,\hat{Y})\} = \int_y \int_z D(y,\hat{y}) p_{Z,Y}(z,y) dz dy.$$

MMSE and Conditional Mean

Since $p_{Z,Y}(z,y) = p_{Y|Z}(y|z)p_Z(z)$, we get

$$E\{D(Y,\hat{Y})\} = \int_y \int_z (y-\hat{y})^2 p_{Y|Z}(y|z)p_Z(z)dzdy$$

$$= \int_z \underbrace{\int_y (y-\hat{y})^2 p_{Y|Z}(y|z)dy}_{I(z)} p_Z(z)dz$$

Note that

- $I(z) \geq 0$ is a function of z only (the dependency on y has been integrated out), and $p_Z(z) \geq 0 \; \forall z$.

Therefore, if we can minimize $I(z)$ for each z, we minimize $E\{D(Y,\hat{Y})\}$.

MMSE and Conditional Mean

For a given value of z we minimize $I(z)$, i.e., we solve $\frac{\partial I(z)}{\partial \hat{y}} = 0$:

$$\frac{\partial}{\partial \hat{y}} \left\{ \int_y (y-\hat{y})^2 \, p_{Y|Z}(y|z) dy \right\} = \int_y \frac{\partial}{\partial \hat{y}} (y-\hat{y})^2 p_{Y|Z}(y|z) dy =$$

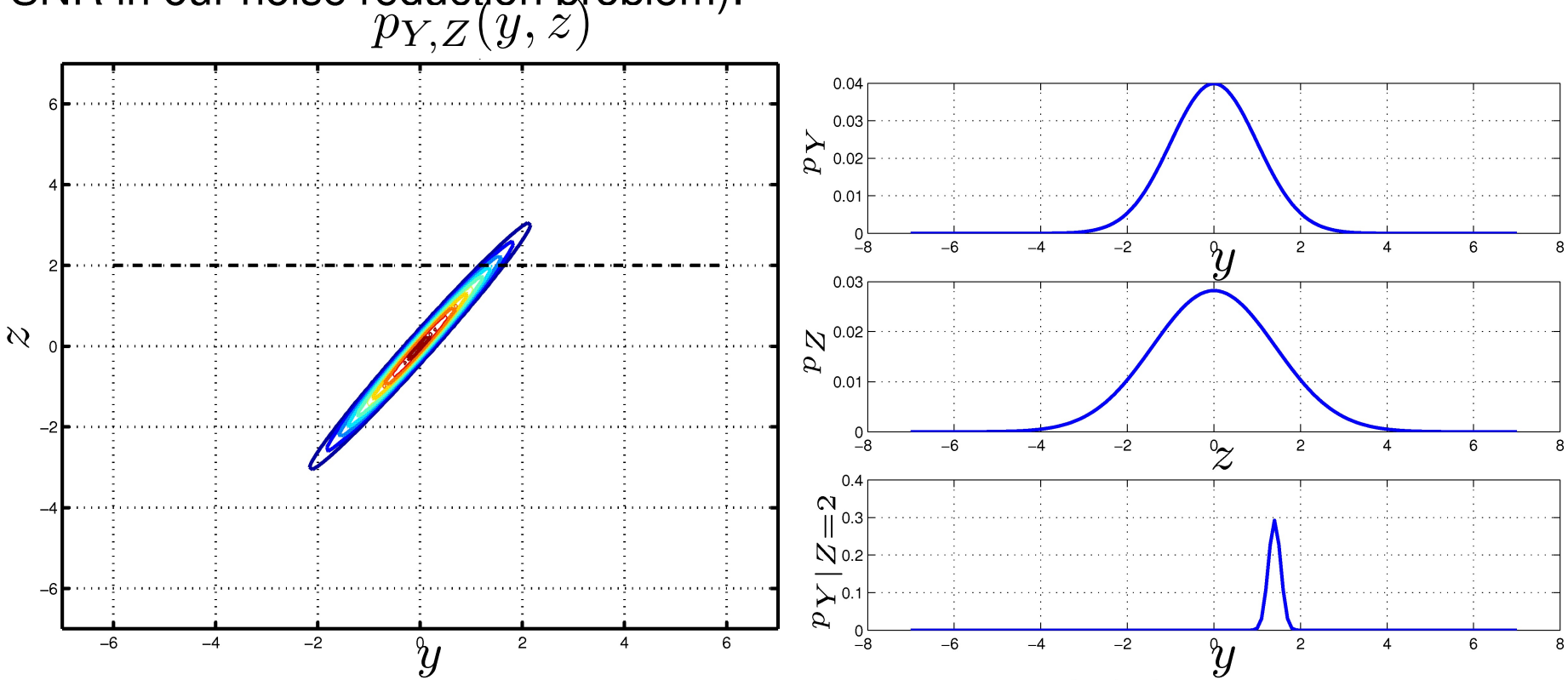
$$-2 \int_y (y-\hat{y}) p_{Y|Z}(y|z) dy = -2 \underbrace{\int_y y p_{Y|Z}(y|z) dy}_{E\{Y|z\}} + 2\hat{y} \underbrace{\int_y p_{Y|Z}(y|z) dy}_{1}$$

Setting to zero gives $\boxed{\hat{y} = E\{Y|z\}.}$

Conclusion: The MMSE estimator $\hat{Y}$ is identical to the *conditional mean* $E\{Y|Z\}$.

MMSE and Conditional Mean

Example: high correlation between random variables Y and Z (high SNR in our noise reduction problem):

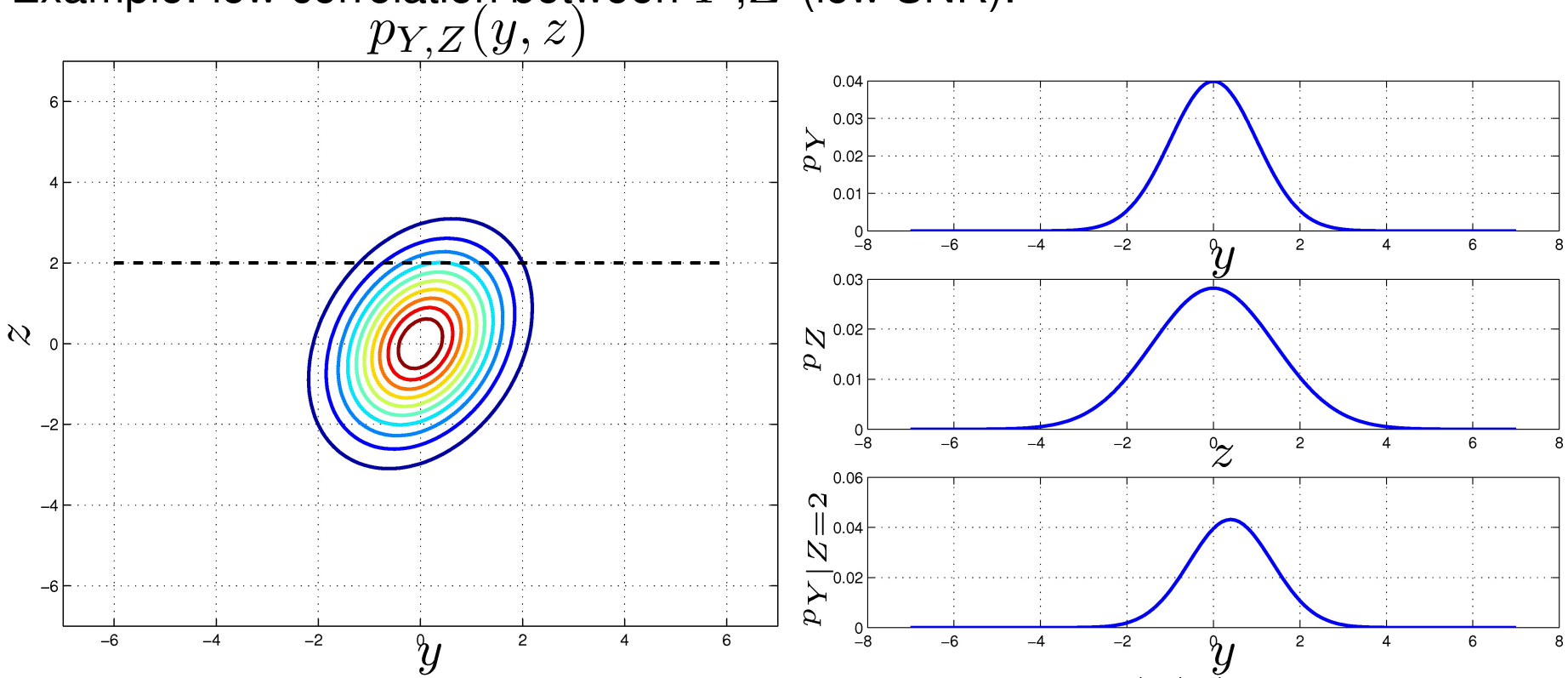


With observation $z = 2$, the posterior density $p_{Y|Z}(y|z)$ is very concentrated around $\hat{y} = E\{Y|z = 2\} \approx 1.5$.

TU Delft

MMSE and Conditional Mean

Example: low correlation between Y,Z (low SNR):



With observation $z = 2$, the posterior density $p_{Y|Z}(y|z)$ is much broader than before. Here $\hat{y} = E\{Y|z\} \approx 0.4$.

TU Delft

MMSE Estimation for Noise Reduction

Now we apply the derived theory to our noise suppression problem in the frequency domain. Our signal model is

$$Y_k(l) = S_k(l) + N_k(l).$$

In literature two different classes of estimators can be found.

- Complex-DFT estimators that estimate $S_k(l)$ by $\hat{S}_k(l)$.
- Magnitude-DFT estimators that estimate $A_k(l) = |S_k(l)|$ and construct $\hat{S}_k(l)$ as $\hat{S}_k(l) = \hat{A}_k(l)e^{j\angle Y_k(l)}$.
  (Motivated by the idea that phase is less important.)

MMSE Estimation for Noise Reduction

We can now restate our goal of finding $\hat{S}_k(l)$ by either

the complex-DFT estimator

$$\hat{S}_k(l) = \arg\min_{\hat{S}_k(l)} E\{|S_k(l) - \hat{S}_k(l)|^2\} = E\{S_k(l)|Y_k(l)\}.$$

or the magnitude-DFT estimator

$$\hat{S}_k(l) = \left(\arg\min_{\hat{A}_k(l)} E\{|A_k(l) - \hat{A}_k(l)|^2\} \right) e^{j\angle Y_k(l)} = E\{A_k(l)|Y_k(l)\} e^{j\angle Y_k(l)}$$

In this course we focus on the complex-DFT estimators.

MMSE Estimation for Noise Reduction

Bayes:

$$p_{S|Y}(s_\Re, s_\Im|y) = \frac{p_{Y|S}(y|s_\Re, s_\Im)\; p_S(s_\Re, s_\Im)}{p_Y(y)}$$

$$= \frac{p_{Y|S}(y|s_\Re, s_\Im)\; p_S(s_\Re, s_\Im)}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{Y|S}(y|s_\Re, s_\Im)\; p_S(s_\Re, s_\Im)\, ds_\Re ds_\Im}$$

$$S = s_\Re + j s_\Im$$

For $\mathrm{E}(S|y)$ we get

$$\mathrm{E}(S|y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (s_\Re + j s_\Im)\, p_{S|Y}(s_\Re, s_\Im|y)\, ds_\Re ds_\Im$$

$$= \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (s_\Re + j s_\Im)\, p_{Y|S}(y|s_\Re, s_\Im)\; p_S(s_\Re, s_\Im)\, ds_\Re ds_\Im}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{Y|S}(y|s_\Re, s_\Im)\; p_S(s_\Re, s_\Im)\, ds_\Re ds_\Im},$$

MMSE Estimation for Noise Reduction

In order to compute $\mathrm{E}(S|y)$ we need knowledge of some statistics:

- What is the distribution $p_{Y|S}(y|s_{\Re}, s_{\Im})$?
- What is the distribution $p_S(s_{\Re}, s_{\Im})$ (or its polar transformation $p_{A,\Phi}(a, \phi)$)?
- What can we say about the dependence of A and Φ? If independent, $p_{A,\Phi}(a, \phi) = p_A(a)\ p_\Phi(\phi)$.
- What can we say about the dependence of $S_{\Re}$ and $S_{\Im}$? If independent, $p_S(s_{\Re}, s_{\Im}) = p_{S_{\Re}}(s_{\Re})\ p_{S_{\Im}}(s_{\Im})$.

The distribution of DFT coefficients and the Central Limit Theorem

Through time, speech and noise DFT coefficients have often been assumed complex-Gaussian dsitributed. Why?

- Simplicity
- Central limit theorem:

> Let $X_1, ..., X_n$ be n mutually *independent* random variables with variances $\sigma_1^2, ..., \sigma_n^2$.
> The random variable $Z = X_1 + ... + X_n$ is then Gaussian distributed when it holds that $\sigma_i^2 < \epsilon\sigma_Z^2$ with $i \in \{1, ..., n\}$

The distribution of DFT coefficients and the Central Limit Theorem

Let the time-domain samples have some some distribution, and consider one speech DFT coefficient:

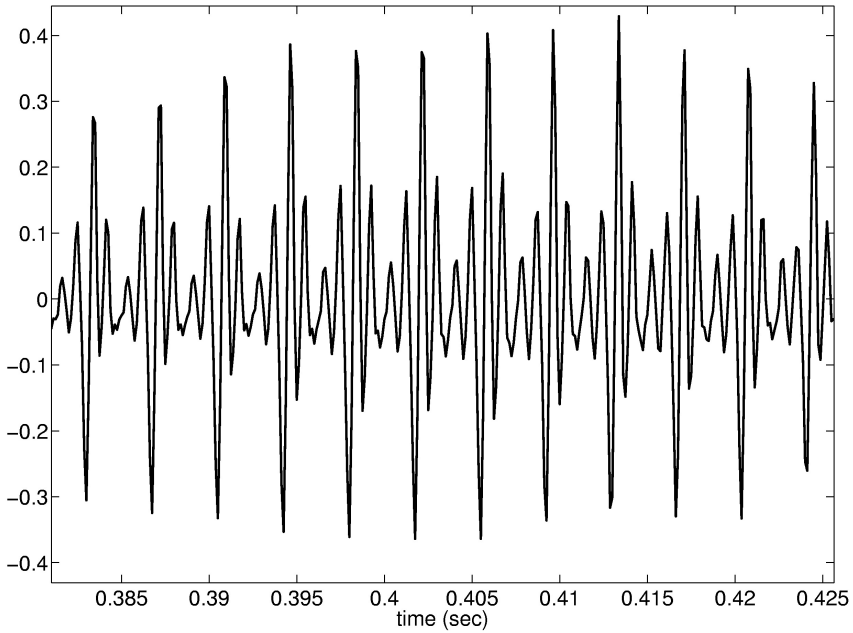
$$S(k) = \sum_{n=0}^{N-1} S_t(n) e^{-\frac{2\pi j}{N} kn}$$

We see that $S(k)$ is a sum of (scaled) random variables.

Is $S(k)$ therefore Gaussian distributed?

Speech DFT coefficients:

Well, in short-time frames, the time samples of a speech signal are not independent:



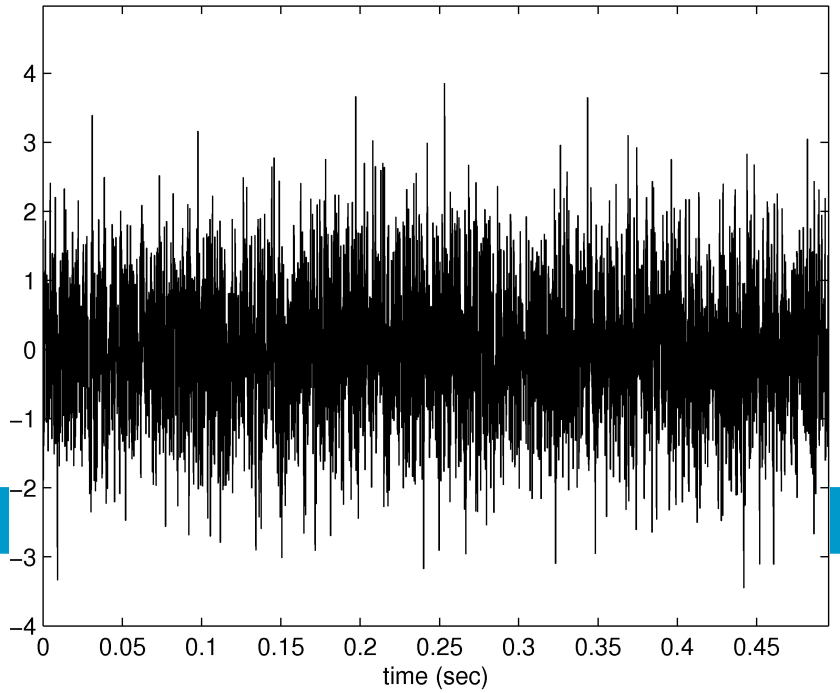
Therefore, the central limit theorem does not apply on short-time speech frames.

Noise DFT coefficients:

For noise DFT coefficients it holds that the time-span of depedency (the time-span over which time samples are still dependent) is rather short.

Therefore, for noise DFT coefficients the central limit theorem applies better and noise DFT coefficients can be considered to be Gaussian distributed.

$$p_N(N) = \frac{1}{\pi\sigma_N^2}\exp\left[-\frac{|N|^2}{\sigma_N^2}\right]$$



TU Delft

Histograms of Noise DFT coefficients:

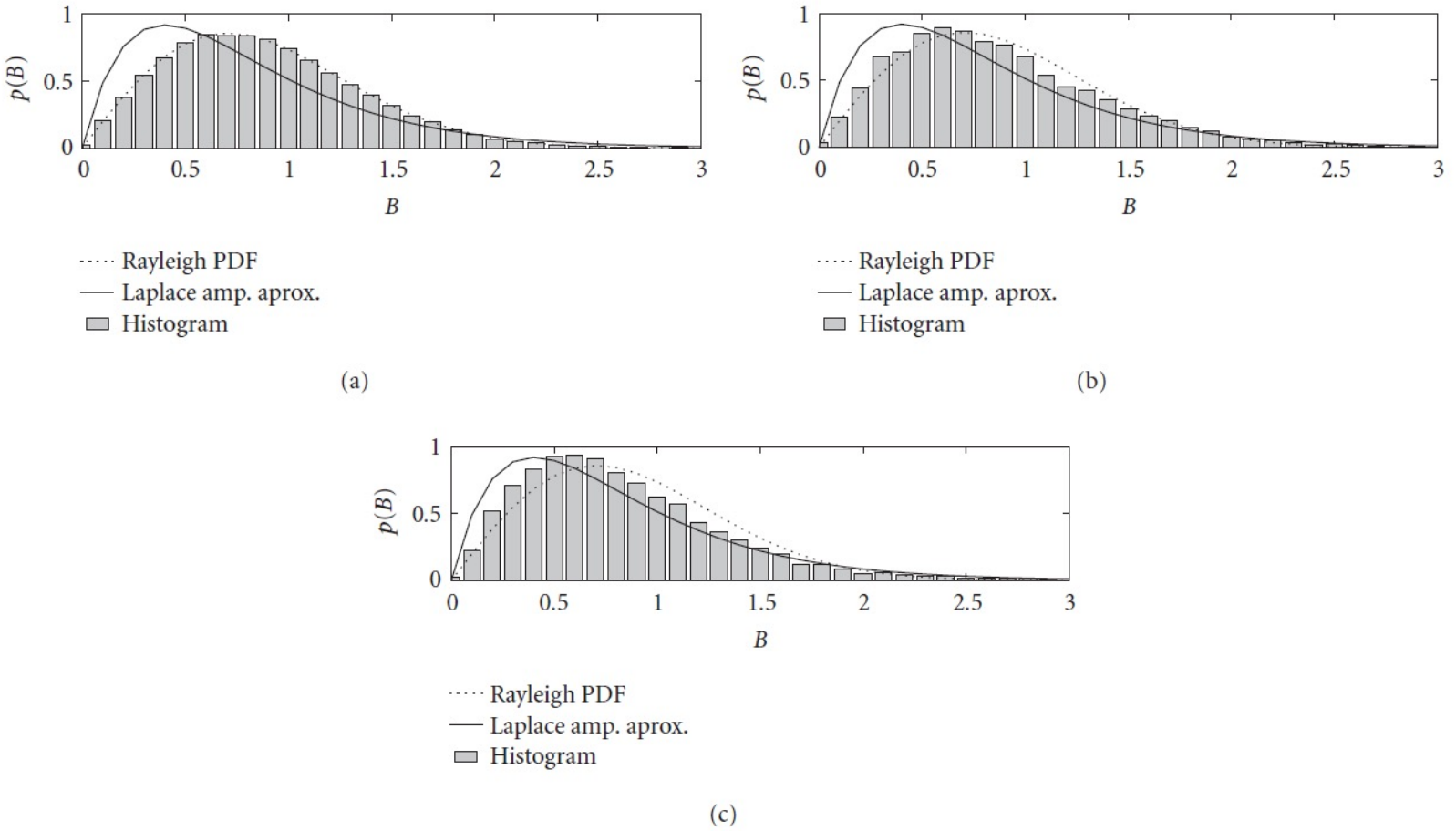


FIGURE 10: Histogram of noise DFT amplitudes B for (a) white uniform distributed noise, (b) fan noise, and (c) cafeteria noise ($\sigma_N^2 = 1$) fitted with Rayleigh PDF and Laplace amplitude approximation.

Taken from: T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model". Eurasip Journal on applied signal processing 2005:7, 1110-1126

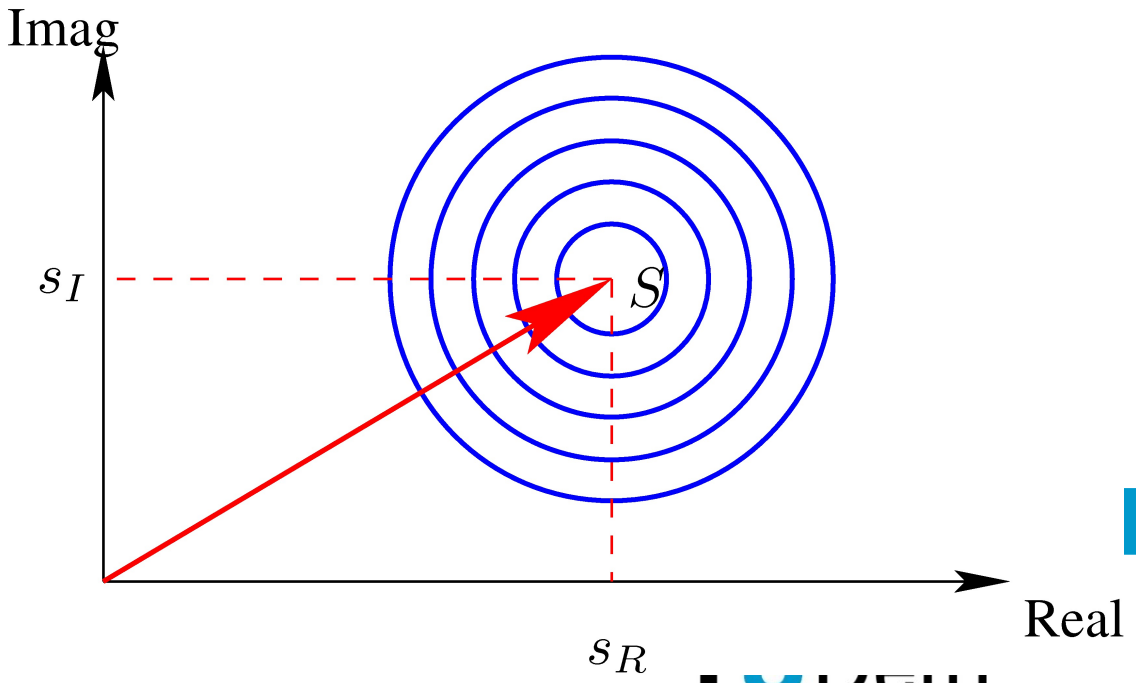
Distributional Assumptions

What does this mean for $p_{Y|S}(y|s)$?

Notice that *given* s, the probability of a certain y is fully determined by the noise n, as $y = s + n$

$$p_{Y|S}(y|s) = \frac{1}{\pi\sigma_N^2}\exp\left(-\frac{1}{\sigma_N^2}|y-s|^2\right)$$

The pdf $p_{Y|S}(y|s)$ is a two-dimensional (complex!) Gaussian with mean s and variance σ_N^2!



MMSE Estimation for Noise Reduction

- What is the distribution $p_{Y|S}(y|s_{\Re}, s_{\Im})$?

$$p_{Y|S}(y|s) = \frac{1}{\pi\sigma_N^2}\exp\left(-\frac{1}{\sigma_N^2}|y-s|^2\right) = p_{Y_{\Re}|S_{\Re}}(y_{\Re}|s_{\Re})\; p_{Y_{\Im}|S_{\Im}}(y_{\Im}|s_{\Im})$$

- What is the distribution $p_S(s_{\Re}, s_{\Im})$ (or its polar transformation $p_{A,\Phi}(a,\phi)$)?
- What can we say about the dependence of A and Φ? If independent, $p_{A,\Phi}(a,\phi) = p_A(a)\; p_{\Phi}(\phi)$.
- What can we say about the dependence of $S_{\Re}$ and $S_{\Im}$? If independent, $p_S(s_{\Re}, s_{\Im}) = p_{S_{\Re}}(s_{\Re})\; p_{S_{\Im}}(s_{\Im})$.

Are Real and imaginary parts of DFT coefficients independent?

What about the distribution $p_S(s_\Re, s_\Im)$?

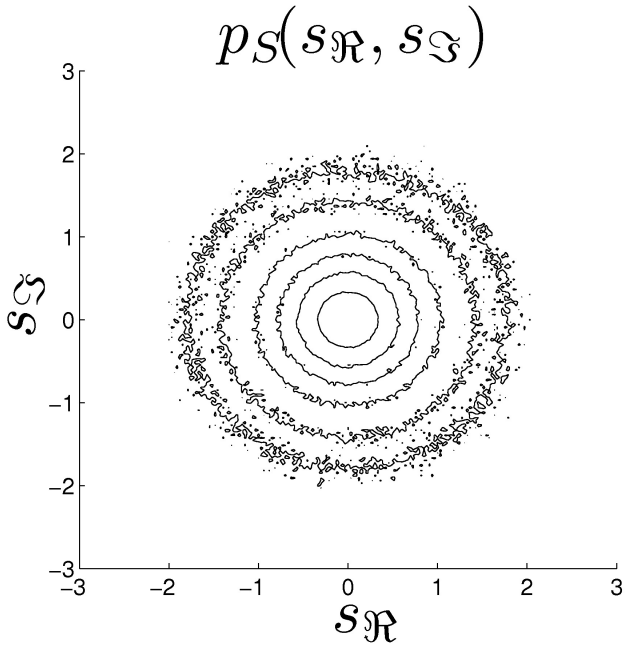
If $S_\Re$ and $S_\Im$ would be independent, $p_S(s_\Re, s_\Im)$ could be written as $p_{S_\Re}(s_\Re)\ p_{S_\Im}(s_\Im)$.

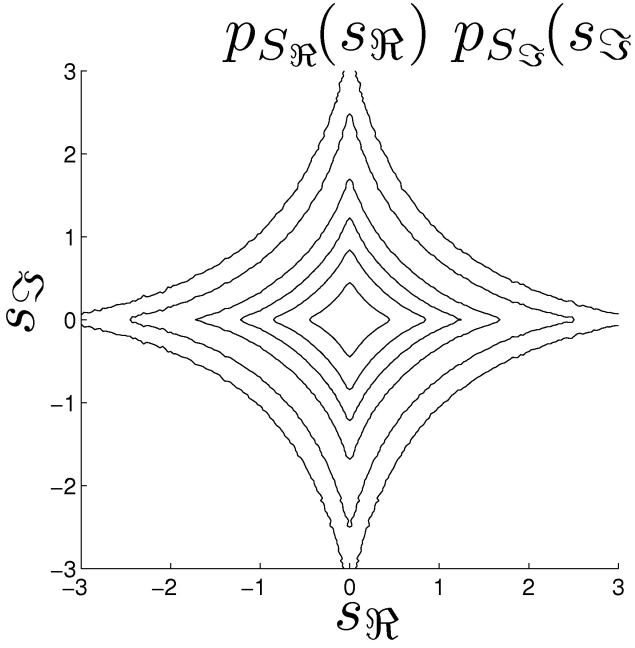
If this holds,

$$E[S|Y] \quad = \quad E[S_\Re|Y_\Re] + jE[S_\Im|Y_\Im].$$

Let us check whether $S_\Re$ and $S_\Im$ are indeed independent.

Distribution of speech DFTs





Observations:

- $p_S(s_\Re, s_\Im) \neq p_{S_\Re}(s_\Re)\ p_{S_\Im}(s_\Im) \Rightarrow S_\Re$ and $S_\Im$ NOT statistically independent
- $p_S(s_\Re, s_\Im)$ is circulair symmetric $\Rightarrow$
  Write $S = Ae^{j\Phi}$: phase Φ is uniform and independent from A.
  $p_\Phi(\phi) = \frac{1}{2\pi}$

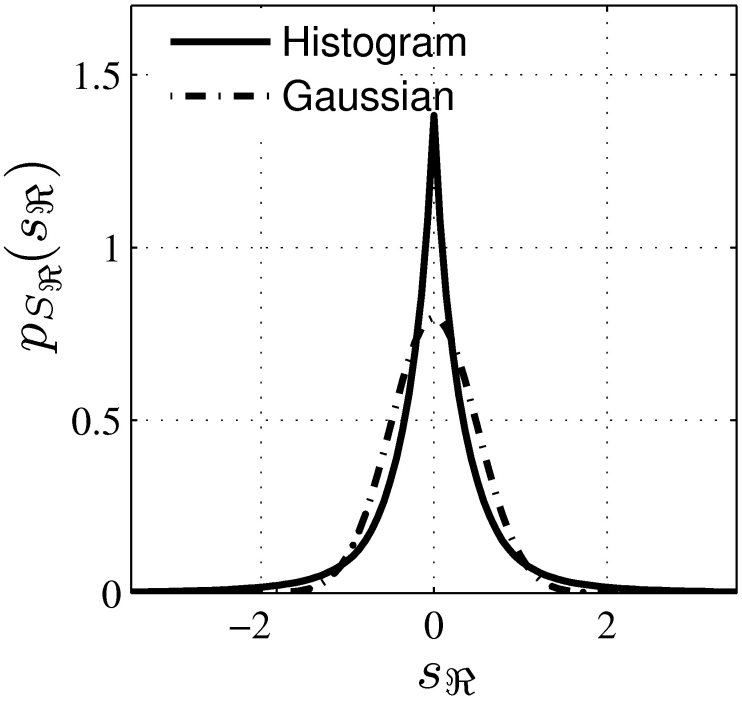
Distribution of speech DFTs

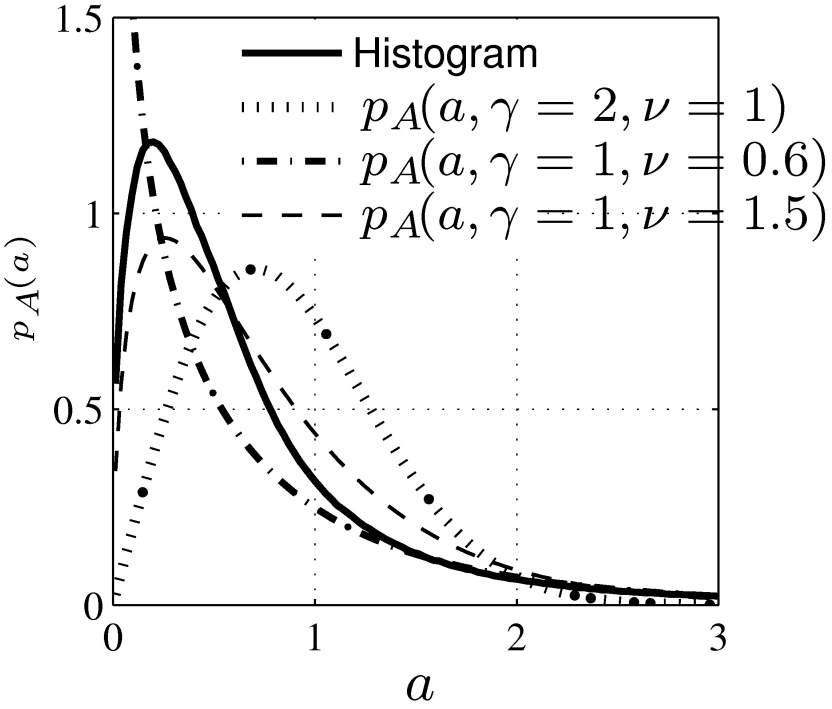
Conclusion:

$$p_S(s_{\Re}, s_{\Im}) \neq p_{S_{\Re}}(s_{\Re}) p_{S_{\Im}}(s_{\Im})$$

$$p_S(s_{\Re}, s_{\Im}) = \frac{1}{a} p_{A,\Phi}(a, \phi) = \frac{1}{a} p_A(a) p_{\Phi}(\phi) = \frac{1}{a} p_A(a) \frac{1}{2\pi}$$

Distribution of speech DFTs

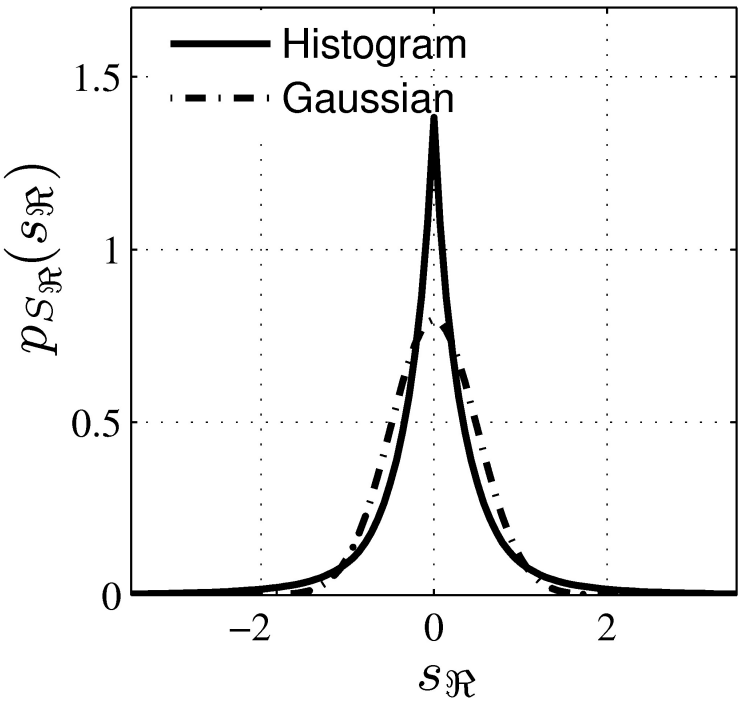


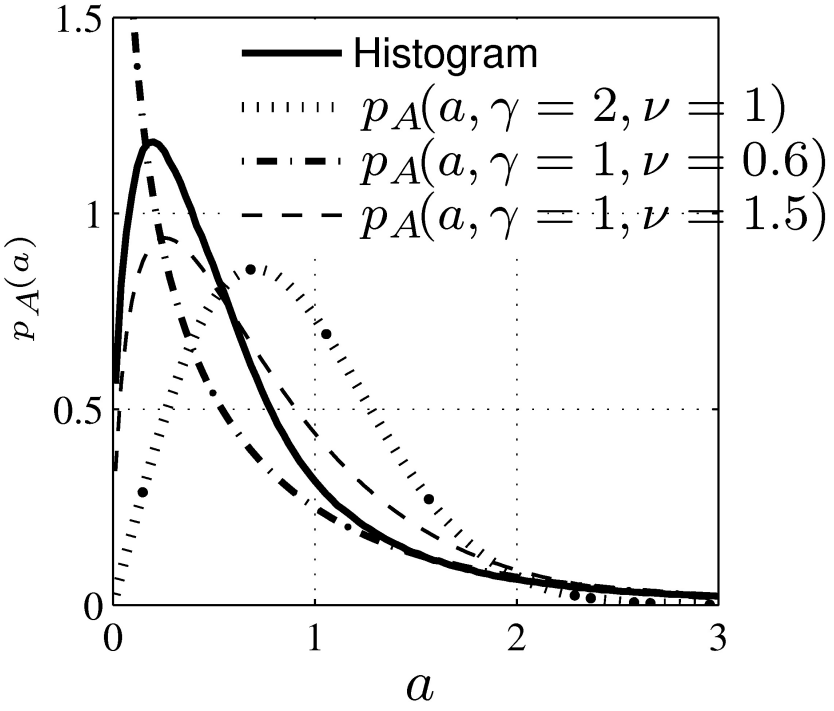


Apparently, speech DFTs are not Gaussian distributed.
Instead, such peaked distributions are known as super-Gaussian distributions.

Distribution of speech DFTs

Rayleigh





$p_{A,\Phi}(a,\phi) = p_A(a)\, p_\Phi(\phi) = p_A(a)\, \frac{1}{2\pi}$ with parametric description for $p_A(a)$

$$p_A(a;\gamma,\nu) = \frac{\gamma\beta^\nu}{\Gamma(\nu)} a^{\gamma\nu-1}\exp\left(-\beta a^\gamma\right),\ \gamma > 0, \nu > 0.$$

TU Delft

Distributional Assumptions – Conclusion:

- Distribution $p_{Y|S}(y|s_{\Re}, s_{\Im})$:

$$p_{Y|S}(y|s) = \frac{1}{\pi\sigma_N^2}\exp\left(-\frac{1}{\sigma_N^2}|y-s|^2\right) = p_{Y_{\Re}|S_{\Re}}(y_{\Re}|s_{\Re})\; p_{Y_{\Im}|S_{\Im}}(y_{\Im}|s_{\Im})$$

- Are A and Φ independent: $p_{A,\Phi}(a,\phi) = p_A(a)\; p_\Phi(\phi)$.
- Distribution of Φ uniform: $p_\Phi(\phi) = \frac{1}{2\pi}$
- Distribution $p_A(a)$:

$$p_A(a;\gamma,\nu) = \frac{\gamma\beta^\nu}{\Gamma(\nu)}a^{\gamma\nu-1}\exp\left(-\beta a^\gamma\right),\ \gamma>0, \nu>0.$$

The Estimator E[S|y]

$$E\left[S|Y\right] = \quad E\left[S_{\Re}|Y_{\Re}\right] + jE\left[S_{\Im}|Y_{\Im}\right]$$

Does not hold, as $S_{\Re}$ and $S_{\Im}$ are not independent.

The picture can't be displayed.

We know that A and Φ are independent. Let us first tranform the integrals and density $p_S(s_{\Re}, s_{\Im})$ in $E\left[S|Y\right]$ into polar-representation:

$$E\left[S|Y\right] = \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(s_{\Re}+js_{\Im})\,p_{Y|S}(y|s_{\Re},s_{\Im})\;p_S(s_{\Re},s_{\Im})\,ds_{\Re}ds_{\Im}}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}p_{Y|S}(y|s_{\Re},s_{\Im})\;p_S(s_{\Re},s_{\Im})\,ds_{\Re}ds_{\Im}}$$

Jacobians of the two transforms

$$= \frac{\int_{0}^{2\pi}\int_{0}^{+\infty}(s_{\Re}+js_{\Im})\,p_{Y|S}(y|s_{\Re},s_{\Im})\;p_{A,\Phi}(a,\phi)\,\frac{1}{a}adad\phi}{\int_{0}^{2\pi}\int_{0}^{+\infty}p_{Y|S}(y|s_{\Re},s_{\Im})\;p_{A,\Phi}(a,\phi)\,\frac{1}{a}adad\phi}$$

The Estimator E[S|y]

use $p_{A,\Phi}(a,\phi) = p_A(a)\,\frac{1}{2\pi} = p_A(a)\,\frac{1}{2\pi}$

$$E[S|y] = \frac{\int_0^{2\pi}\int_0^{+\infty}(s_\Re + j s_\Im)\, p_{Y|S}(y|s_\Re, s_\Im)\; p_{A,\Phi}(a,\phi)\, da d\phi}{\int_0^{2\pi}\int_0^{+\infty} p_{Y|S}(y|s_\Re, s_\Im)\; p_{A,\Phi}(a,\phi)\, da d\phi}$$

$$= \frac{\int_0^{2\pi}\int_0^{+\infty}(s_\Re + j s_\Im)\, p_{Y|S}(y|s_\Re, s_\Im)\; p_A(a)\,\frac{1}{2\pi} da d\phi}{\int_0^{2\pi}\int_0^{+\infty} p_{Y|S}(y|s_\Re, s_\Im)\; p_A(a)\,\frac{1}{2\pi} da d\phi}$$

$$= \frac{\int_0^{2\pi}\int_0^{+\infty}(s_\Re + j s_\Im)\, p_{Y|S}(y|s_\Re, s_\Im)\; p_A(a)\, da d\phi}{\int_0^{2\pi}\int_0^{+\infty} p_{Y|S}(y|s_\Re, s_\Im)\; p_A(a)\, da d\phi}$$

known known

Substitute the two densities and solve the integrals.

The Estimator E[S|y]

Computing the estimator leads under the right assumptions leads to a function

$$E\left[S|y\right] \quad = \quad g\left(\sigma_N^2, \sigma_S^2, y, \nu, \gamma\right) y$$

Properties of the gain function $g\left(\sigma_N^2, \sigma_S^2, y, \nu, \gamma\right)$:

- $g\left(\sigma_N^2, \sigma_S^2, y, \nu, \gamma\right)$ is a real function $\Rightarrow$ Phase of $E\left[S|y\right]$ is noisy phase!
- $g\left(\sigma_N^2, \sigma_S^2, y, \nu, \gamma\right) > 0$

$$p_A(a;\gamma,\nu) = \frac{\gamma\beta^\nu}{\Gamma(\nu)} a^{\gamma\nu-1}\exp\left(-\beta a^\gamma\right)$$

Hence: Setting $\nu = 1$ and $\gamma = 2$ in $p_A(a)$ gives Rayleigh density for magnitudes (and thus a Complex-Gaussian for the complex-DFTs). Using these parameters for $p_A(a)$, the gain $g\left(\sigma_N^2, \sigma_S^2, y, \nu, \gamma\right)$ leads to

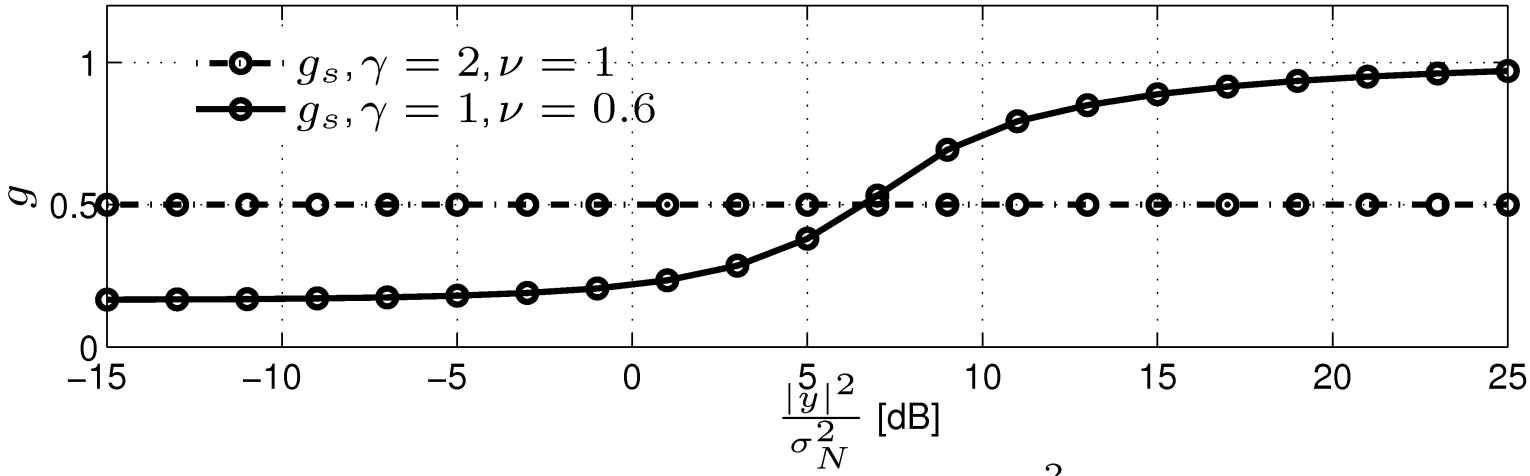
$$g\left(\sigma_N^2, \sigma_S^2, y, \nu, \gamma\right) = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_N^2}$$

Wiener!

Examples of Gain Functions

Super-Gaussian based gain

Wiener gain



Examples of gain curves $g\left(\sigma_N^2, \sigma_S^2, y, \nu, \gamma\right)$ for $\frac{\sigma_S^2}{\sigma_N^2} = 0$ dB.

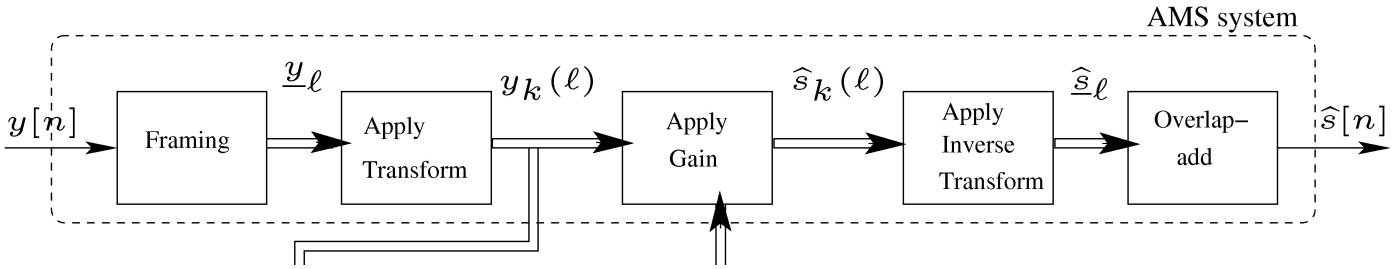
TU Delft

MMSE Estimation for Noise Reduction

Conclusions:

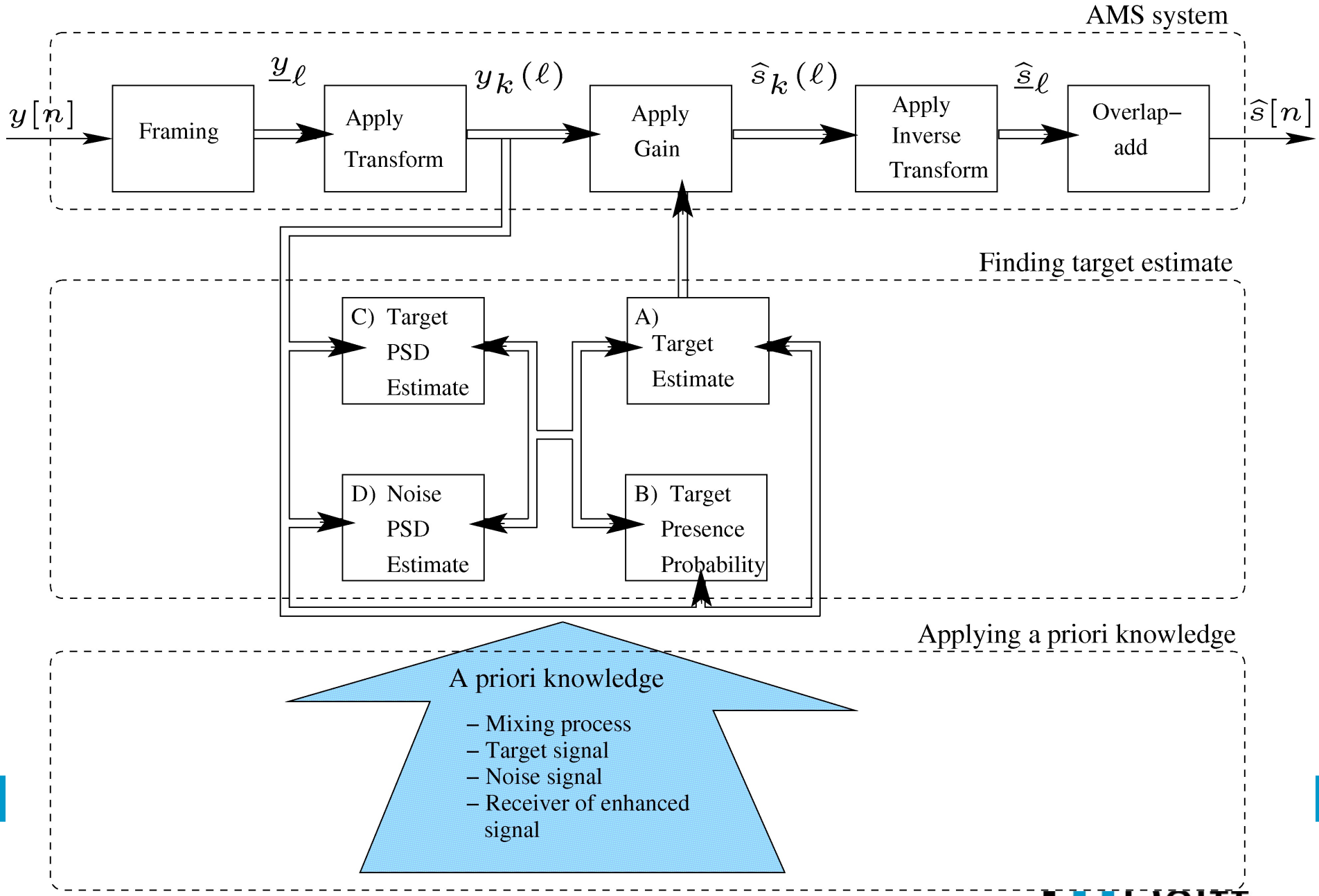
- Wiener gain is the optimal LINEAR estimator.
- The optimal estimator is $E\left[S|y\right] = g\left(\sigma_N^2, \sigma_S^2, y, \nu, \gamma\right) y$ and is generally non-linear with respect to y.
- for Rayleigh $p_A(a)$ (and thus a Complex-Gaussian pdf for the complex-DFTs) Wiener gain results.
- Noisy phase is always used (and is optimal.)

Demonstration



- Noisy 5 dB SNR
- Wiener filter complex DFTs - E[S|y]
- Super-Gaussian complex DFTs - E[S|y]
- Gaussian magnitude DFTs - E[A|y]
- Super-Gaussian magnitude DFTs - E[A|y]

Overview of single-channel NR algorithm



A priori SNR estimation

We see that the suppression rules derived so far rely on the quantity

$$\xi = \frac{\sigma^2_{S,k}(l)}{\sigma^2_{N,k}(l)} = \frac{P_{SS,k}}{P_{NN,k}} = \frac{E\{|S_k(l)|^2\}}{E\{|N_k(l)|^2\}}.$$

which is known in literature as the *a priori SNR*. This key quantity is defined in terms of *expected values*. Estimation of σ^2_N will be discussed next week. But how to estimate σ^2_S?

A priori SNR estimation

'The Maximum-likelihood approach':

We can write ξ in frame l at frequency bin k as

$$\xi_k(l) = \frac{E\{|S_k(l)|^2\}}{E\{|N_k(l)|^2\}} = \frac{E\{|Y_k(l)|^2\}}{E\{|N_k(l)|^2\}} - 1.$$

Obviously, $\xi_k(l)$ can be estimated using a Bartlett estimate of $E\{|Y_k(l)|^2\}$:

$$\hat{\xi}_k(l) = \frac{\frac{1}{K}\sum_{m=l-K+1}^{m} \hat{P}_{YY,k}(m)}{\frac{1}{L}E\{|N_k(l)|^2\}} - 1,$$

where $\hat{P}_{YY,k}(m) = \frac{1}{L}|Y_k(m)|^2$ is the periodogram estimate in frame m.

A priori SNR estimation – ML approach

Why is this called ”maximum likelihood” approach?

Remember that under Gaussian distributional assumptions the ML estimate of $\sigma^2_{S,k}(l)$ given the distribution $p_Y(y(l); \sigma^2_{S,k}(l), \sigma^2_{N,k}(l))$ for one frame l was given by

$$\widehat{\sigma^2_{S,k}(l)}_{ml} = \arg \max_{\sigma^2_{S,k}(l)} p_Y(y(l); \sigma^2_{S,k}(l), \sigma^2_{N,k}(l)) = |y_k(l)|^2 - \sigma^2_{N,k}(l)$$

The ML estimate given time-frames $m = l - K + 1 ... l$ is then given by maximizing the joint density (assuming $\sigma^2_{S,k}(l)$ and $\sigma^2_{N,k}(l)$ are constant over time):

$$\widehat{\sigma^2_{S,k}}_{ml} = \arg \max_{\sigma^2_{S,k}} \prod_{m=l-K+1}^{l} p_Y(y(m); \sigma^2_{S,k}, \sigma^2_{N,k}) = \frac{1}{K} \sum_{m=l-K+1}^{l} |y_k(m)|^2 - \sigma^2_{N,k}$$

A priori SNR estimation

'The decision-directed (DD) approach':

The DD approach uses the observation

$$\xi_k(l) = \alpha \overbrace{\frac{E\{|S_k(l)|^2\}}{E\{|N_k(l)|^2\}}}^{I} + (1-\alpha) \overbrace{\left(\frac{E\{|Y_k(l)|^2\}}{E\{|N_k(l)|^2\}} - 1 \right)}^{II}.$$

I: To estimate $E\{|S_k(l)^2\}$ we drop the expectation operator. Further, assuming that the power does not change fast across time, we replace $|S_k(l)|^2$ by the estimate from the previous frame $|\hat{S}_k(l-1)|^2$.

A priori SNR estimation

'The decision-directed (DD) approach':

II: To estimate $E\{|Y_k(l)|^2\}$ we drop the expectation operator, i.e., use a periodogram estimator. In doing so, the expression in II may become negative (which does not make sense for a psd estimate). We therefore set negative psd estimates to zero.

We get

$$\hat{\xi}_m(\omega) = \alpha \frac{|\hat{S}_k(l-1)|^2}{E\{|N_k(l)|^2\}} + (1-\alpha)\max\left(\frac{|Y_k(l)|^2}{E\{|N_k(l)|^2\}} - 1{,}0\right).$$

A priori SNR estimation

'The decision-directed (DD) approach':

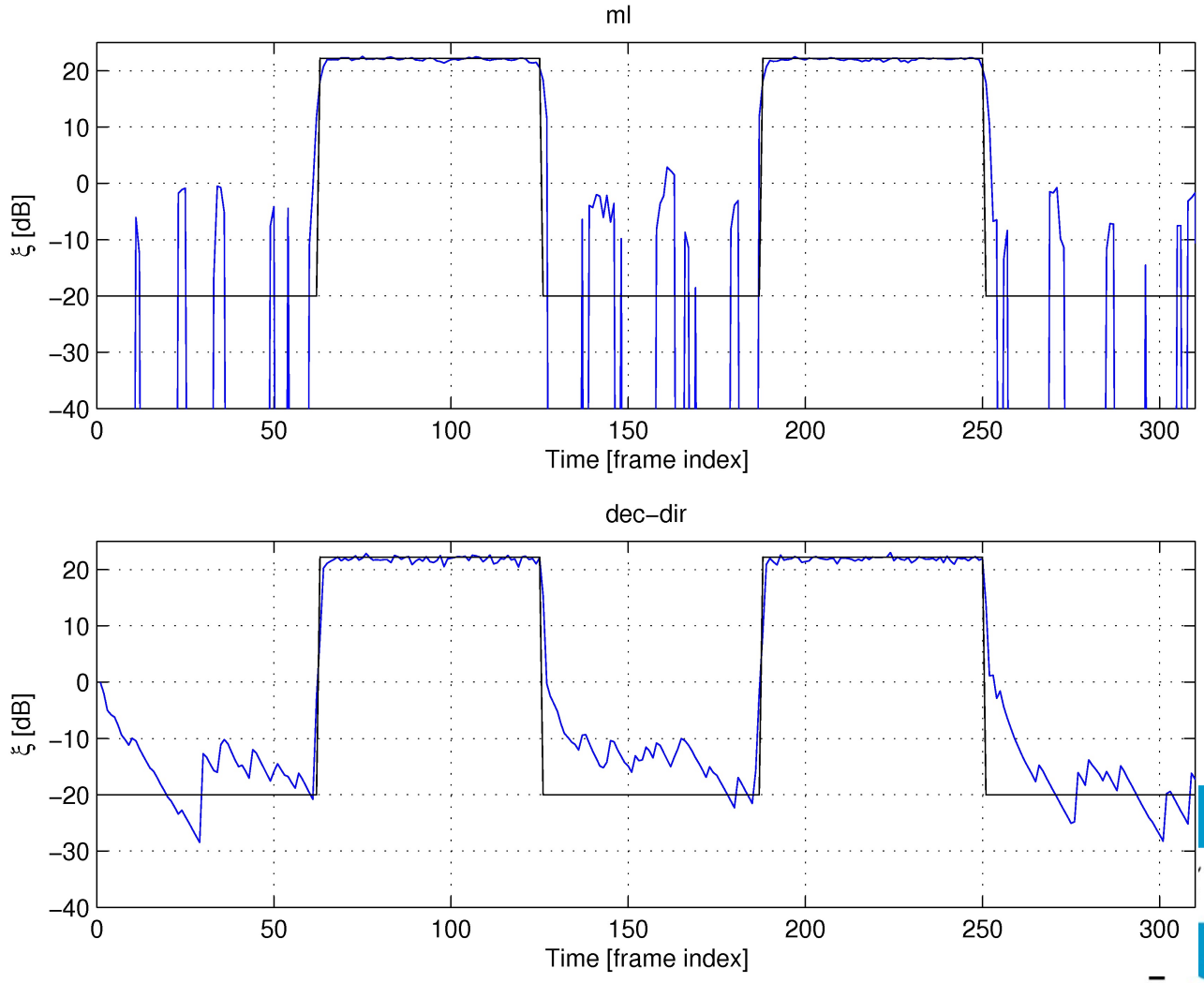
How to choose $0 \geq \alpha \geq 1$?

- $\alpha \to 1$ implies more smoothing but higher delay for tracking rapid changes.
- $\alpha \to 0$: current frame has more weight and tracking delay is reduced at the expense of higher variance in $\hat{\xi}_k(l)$
- Based on objective quality measures and subjective evaluations, α is typically chosen as $\alpha \approx 0.96 - 0.99$.

A priori SNR estimation

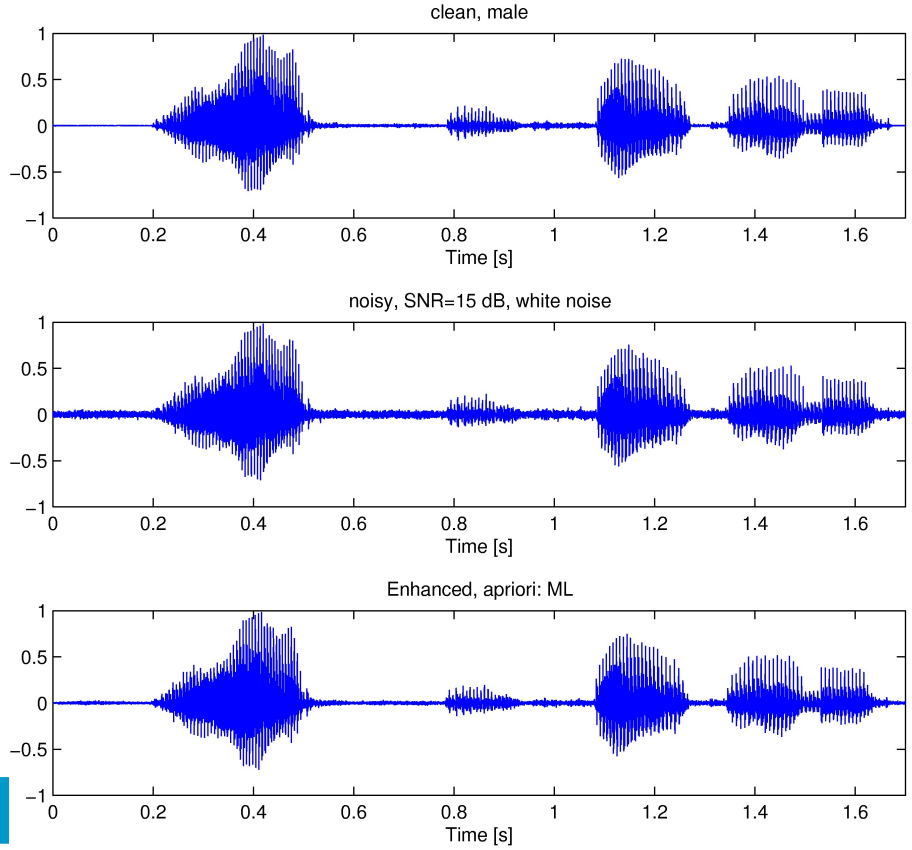
Example: ML vs. Decision Direction ($\alpha = 0.98$).

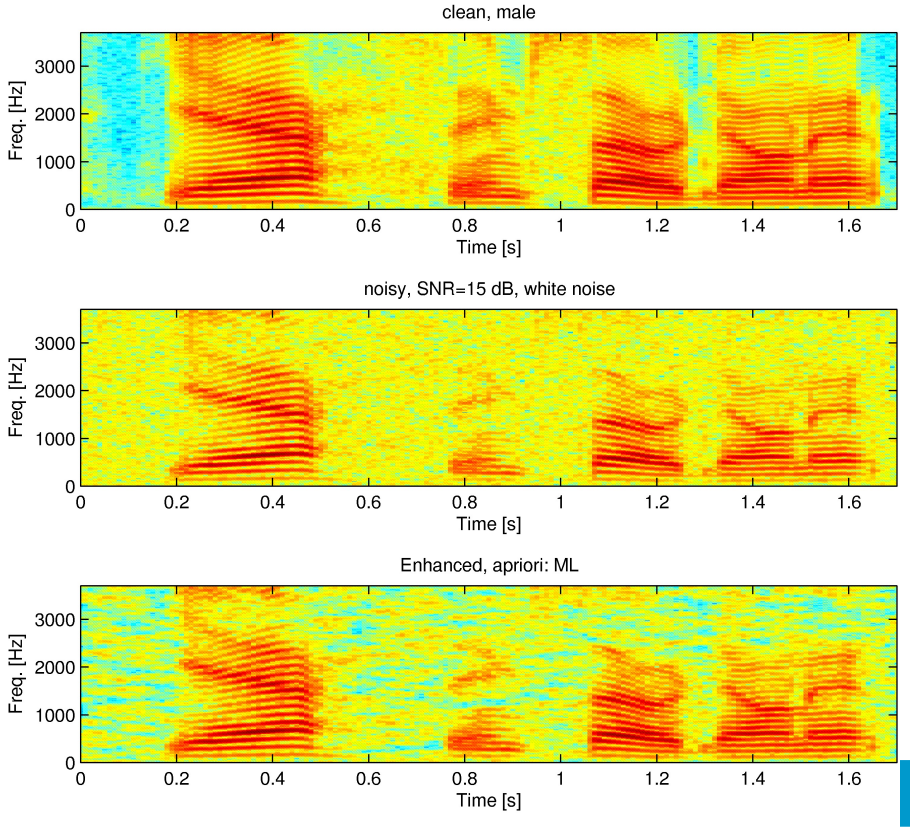
Plots show $\hat{\xi}_k(l)$ for a given bin k for $m = 0, 1, 2, \ldots$.



Enhancement Results

- SNR estimated with ML approach.
- Estimator: Wiener smoother.
- Input SNR: 15 dB (white Gaussian noise).



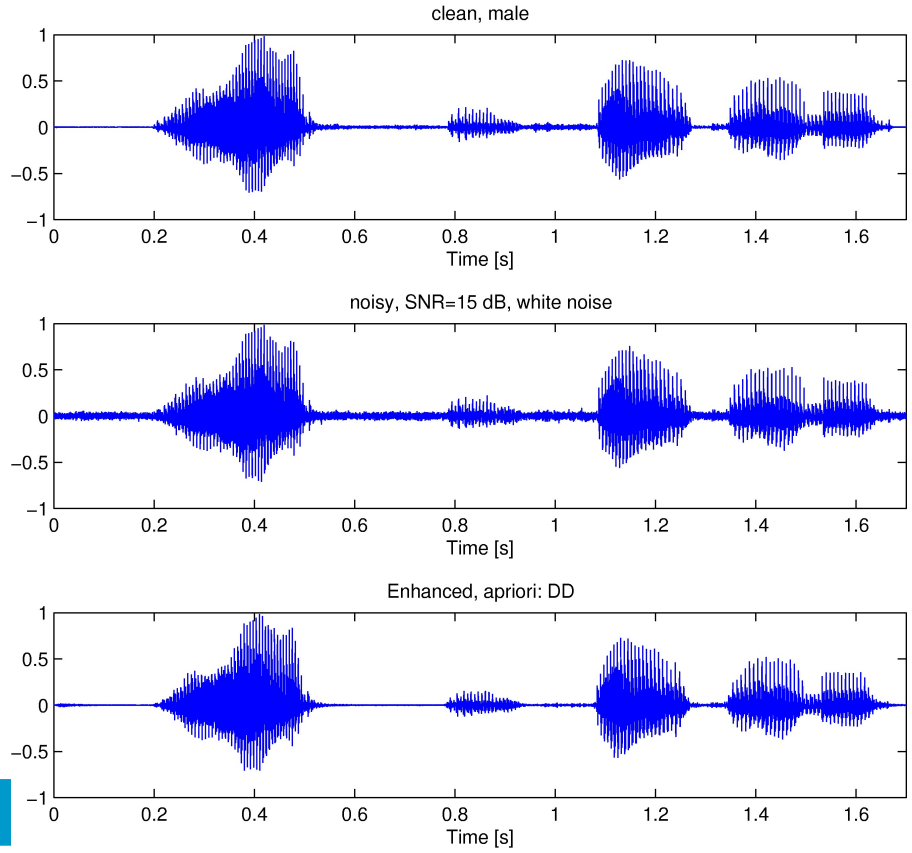


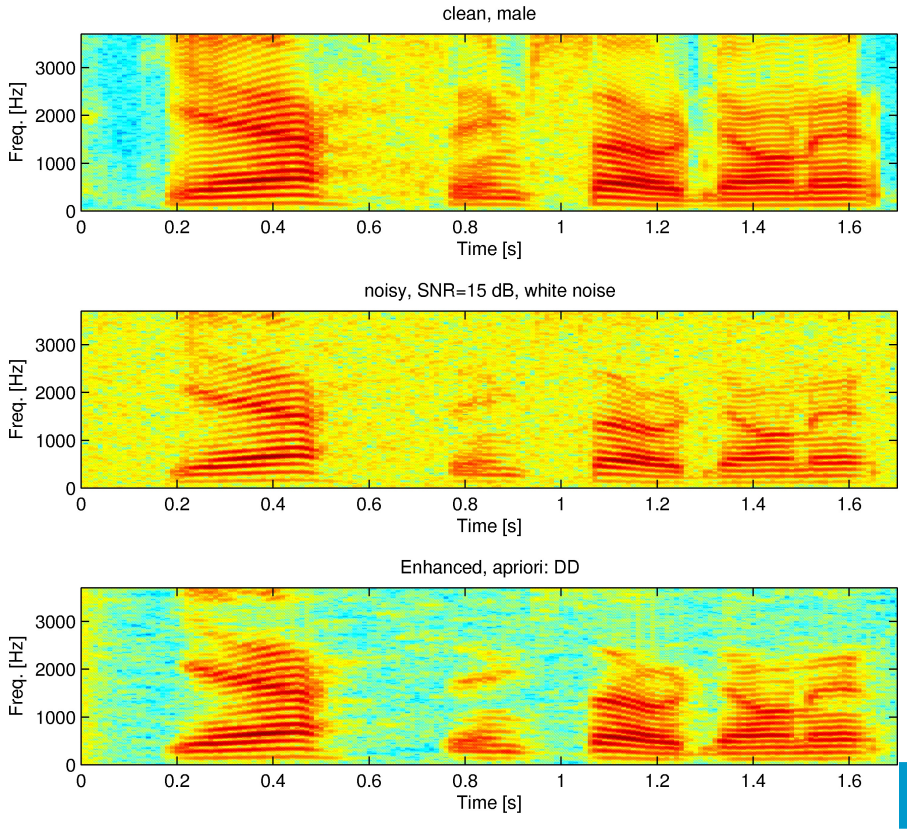
clean noisy Processed ML approach

TU Delft

Enhancement Results

- A priori SNR estimated with DD approach.
- Estimator: Wiener smoother.
- Input SNR: 15 dB (white Gaussian noise).





clean noisy Processed DD approach

TU Delft

Overview of single-channel NR algorithm