Model-Based Speech Enhancement Using SNR-Dependent MMSE Estimation

Outline

- Introduction
- System Overview: Model-Based Noise Reduction
- SNR-Dependent MMSE Estimation
- Evaluation, Test Results, and Demonstration
- Summary

Institute of Communication Systems and Data Processing
Prof. Dr.-Ing. P. Vary

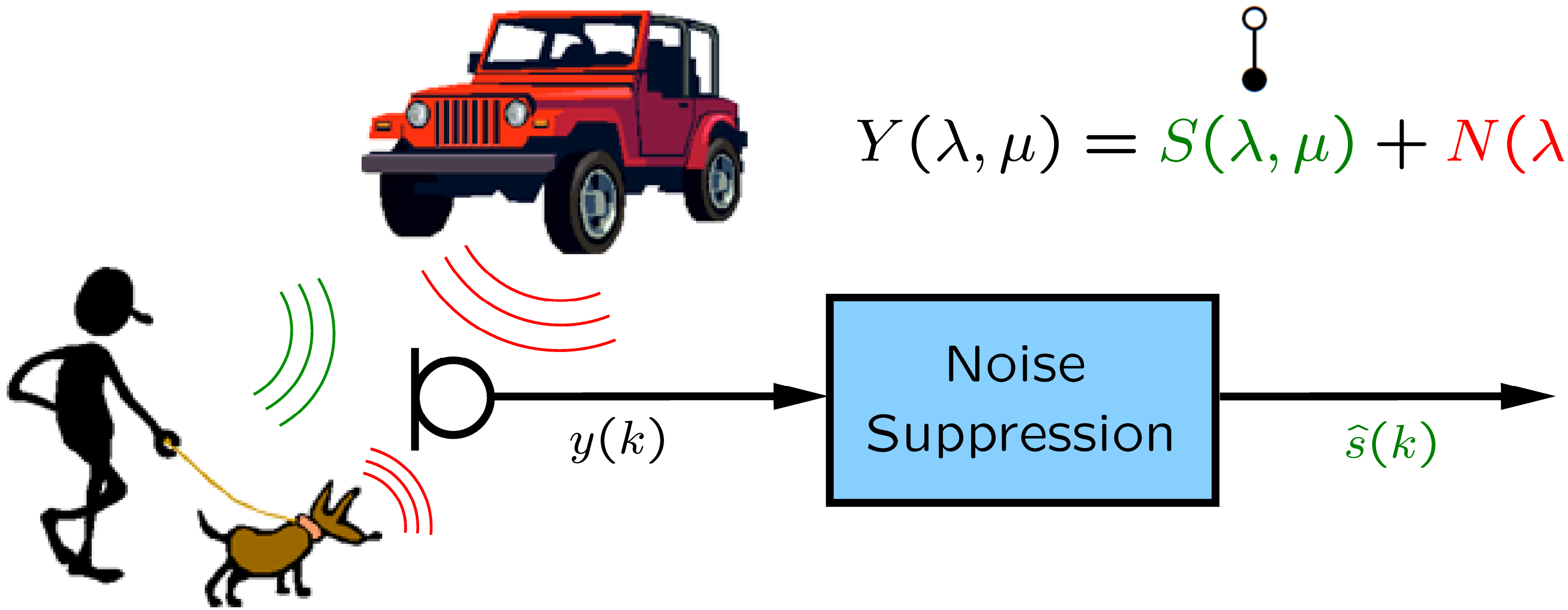
RWTH AACHEN

Introduction

- Ambient noise impairs quality and/or intelligibility of transmitted speech signal in communication device

$$y(k) = s(k) + n(k)$$

$$Y(\lambda, \mu) = S(\lambda, \mu) + N(\lambda, \mu)$$



discrete time index k – frame index λ – frequency index μ

Introduction

► Statistical noise reduction approaches

- Certain assumptions about statistics of speech and noise (e.g., Gaussian or Gamma PDF)
- Mathematical criteria (e.g., MMSE, ML, MAP)

Exploitation of memory-less a priori knowledge

► Model-based approaches

- Consider correlation across time and/or frequency
- Take into account model of speech production system

Exploitation of a priori information of higher order

System Overview: Model-Based Noise Reduction

- Mo… orrelation of
  - …es
  - …eristics

Prediction error of first step:

$$E_S(\lambda,\mu) = S(\lambda,\mu) - \widehat{S}^{\text{prop}}(\lambda,\mu)$$

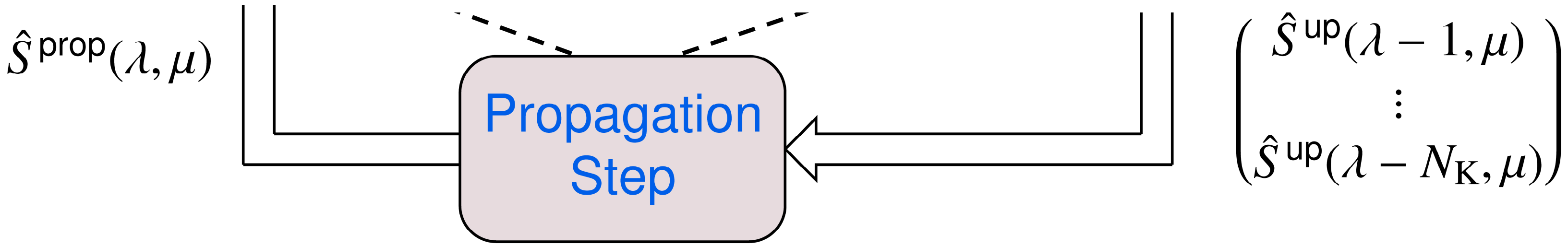
Estimation by spectral weighting of $D(\lambda,\mu)$:

$$\widehat{E}_S(\lambda,\mu) = G(\lambda,\mu) \cdot D(\lambda,\mu)$$

- Separate low-order Kalman filters for each frequency bin

$$\widehat{S}^{\text{prop}}(\lambda,\mu) = \sum_{i=1}^{N_{\text{K}}} \hat{a}_i(\lambda,\mu)\,\widehat{S}^{\text{up}}(\lambda - i,$$

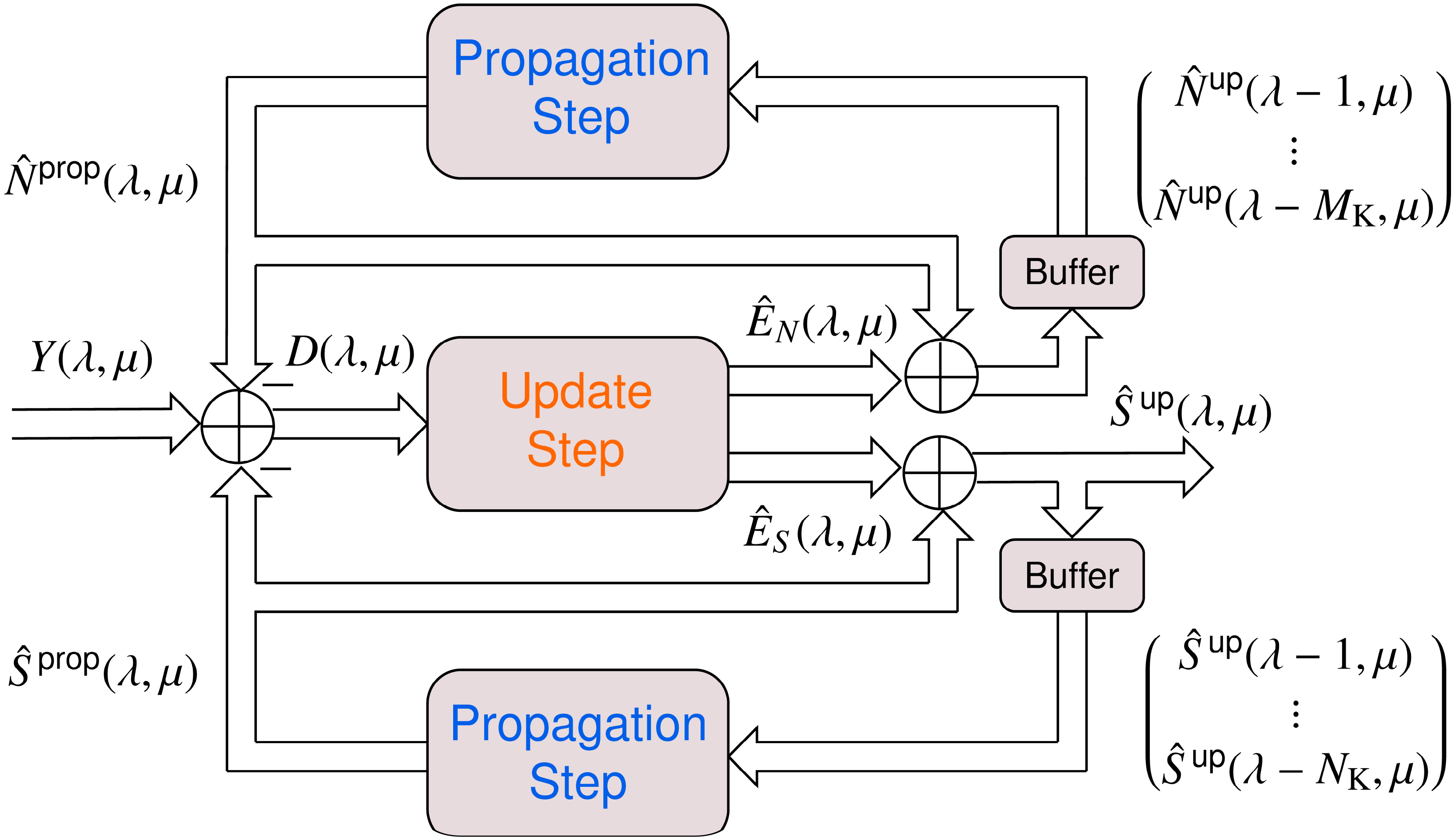
with model order N_{K} and AR coeffi…



frame index λ – frequency index μ

System Overview: Model-Based Noise Reduction

- Extension to noise signals



frame index λ – frequency index μ

System Overview: Model-Based Noise Reduction

- Update Step
  - Objective: Estimate speech and noise prediction errors $E_S(\lambda,\mu)$ and $E_N(\lambda,\mu)$ given differential signal $D(\lambda,\mu)$

$$D(\lambda,\mu) = \underbrace{Y(\lambda,\mu)}_{S(\lambda,\mu)+N(\lambda,\mu)} - \widehat{S}^{\text{prop}}(\lambda,\mu) - \widehat{N}^{\text{prop}}(\lambda,\mu)$$

‘Classical’ noise reduction problem in update step

- Application of conv. estimator adapted to statistics of E_S and E_N

$$\widehat{E}_S(\lambda,\mu) = G(\lambda,\mu)\cdot D(\lambda,\mu) \qquad \widehat{E}_N(\lambda,\mu) = (1-G(\lambda,\mu))\cdot D(\lambda,\mu)$$

SNR-Dependent MMSE Estimation

▶ Derivation of original Kalman filter gain

- Assumption: Gaussian PDF for E_S and E_N
- Minimization of $\mathbb{E}\{|E_S - \widehat{E}_S|^2\}$

Wiener filter solution: $G = \frac{\mathbb{E}\{|E_S|^2\}}{\mathbb{E}\{|E_S|^2\}+\mathbb{E}\{|E_N|^2\}}$

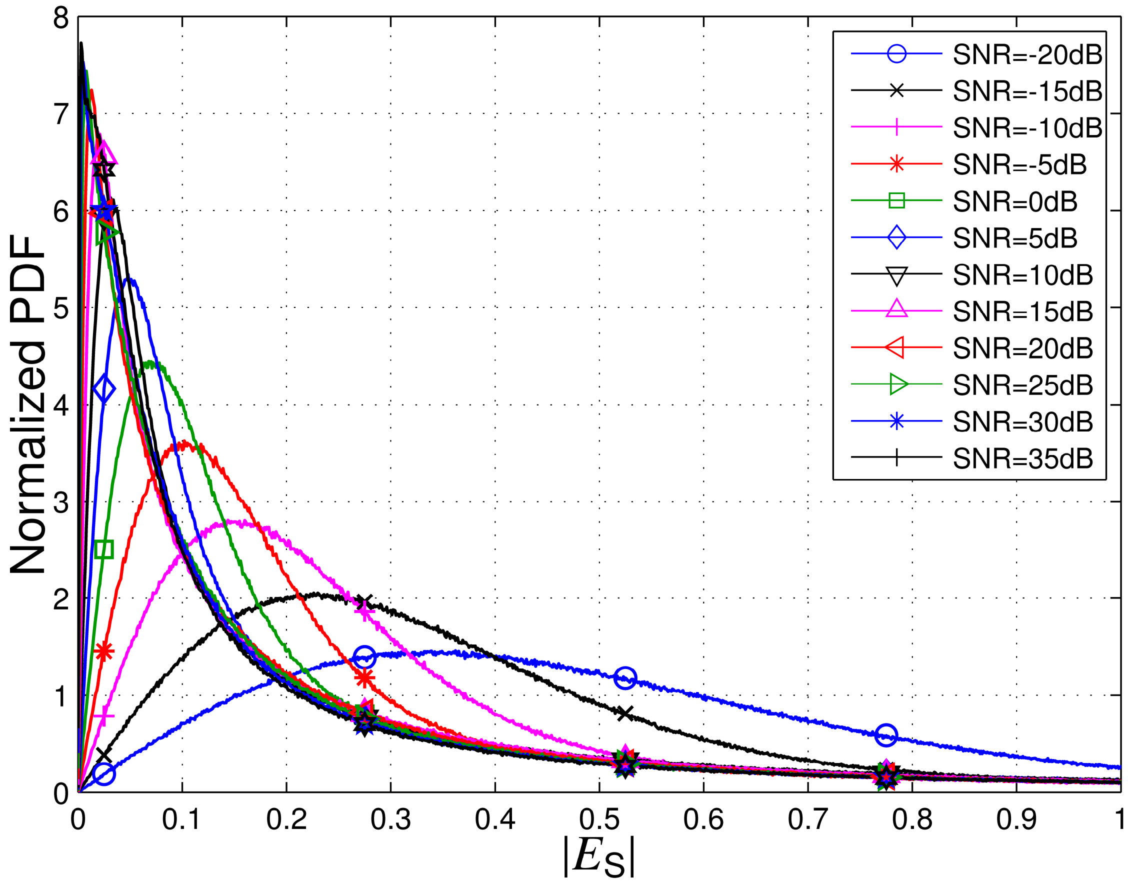
▶ Distribution of E_S is rather super-Gaussian [Esch and Vary, ICASSP 08]

- Can be exploited using adequate statistical estimator, e.g., MMSE estimator under generalized Gamma priors [Erkelens et al., IEEE SigPro. Let. 08]

So far: PDF measurement of E_S averaged over wide SNR range

SNR-Dependent MMSE Estimation

- SNR-dependent histograms for E_S (noise type: WGN)



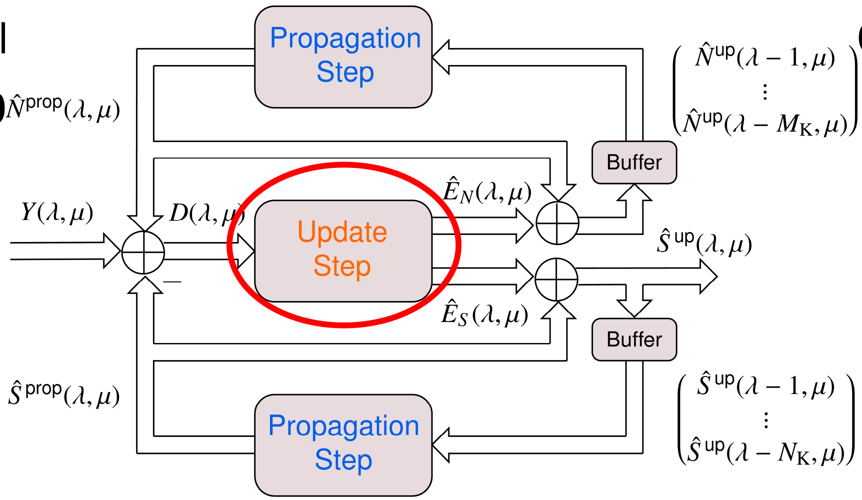
Smaller prediction errors occur proportionally more often at higher input SNR values

SNR-Dependent MMSE Estimation

- Proposed solution: SNR-dependent MMSE estimation in update step
  - Configurable MMSE estimator [Erkelens et al., IEEE SigPro Letters 08] based on generalized Gamma PDF
  - Each SNR value (step size: 5dB) provides *one* parameter set

$$G(\lambda,\mu) = G(\lambda,\mu,\widehat{\mathrm{SNR}})$$

  - SNR estimate $\widehat{\mathrm{SNR}}$ on the basis of enhanced coefficients from previous frames
  - Increase in co ory requirements compared to p

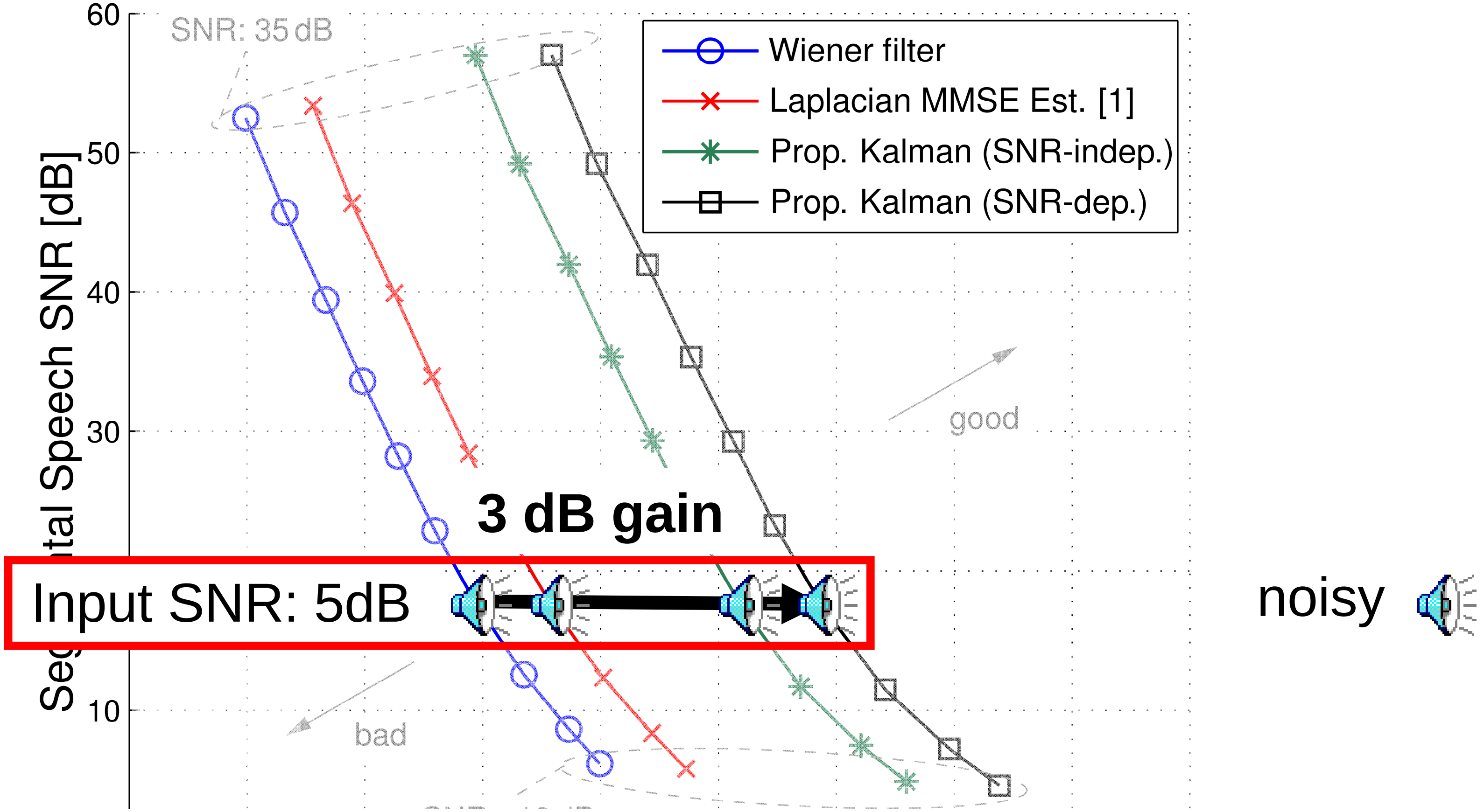


SNR-Dependent MMSE Estimation

- System settings
  - Model orders: $N_{\mathsf{K}} = 3$ (speech) and $M_{\mathsf{K}} = 2$ (noise)
  - AR coefficients estimated in each frame using Levinson-Durbin algorithm applied to estimates from previous frames
  - Minimum Statistics applied in update step for ‘noise’ power estimation

Evaluation, Test Results, and Demonstration

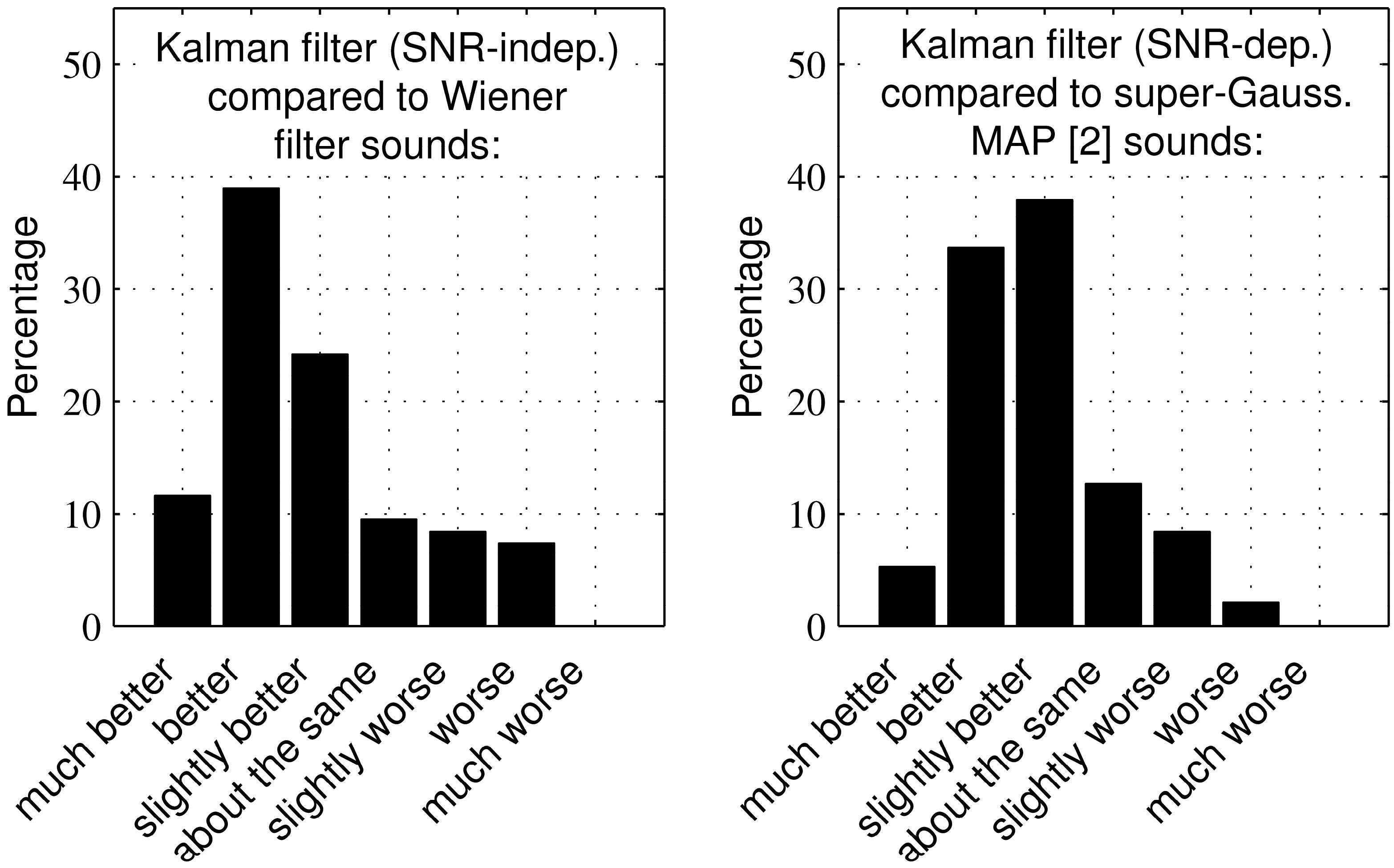
- Objective measurements (f16, babble, car, factory, WGN)



Further objective measurements can be found in the paper.

Evaluation, Test Results, and Demonstration

- Results of informal listening test (19 probands)



[2] Lotter and Vary, EURASIP Journal on Applied Signal Processing 05

Summary

- Modified Kalman filter exploits temporal correlation of speech and noise DFT coefficients
- Input SNR influences statistics of speech prediction error in update step
- Application of SNR-dependent MMSE estimator adapted to measured histograms of speech prediction error signal
- Objective and subjective evaluations show consistent improvements compared to purely statistical estimators

Thank You!

System settings

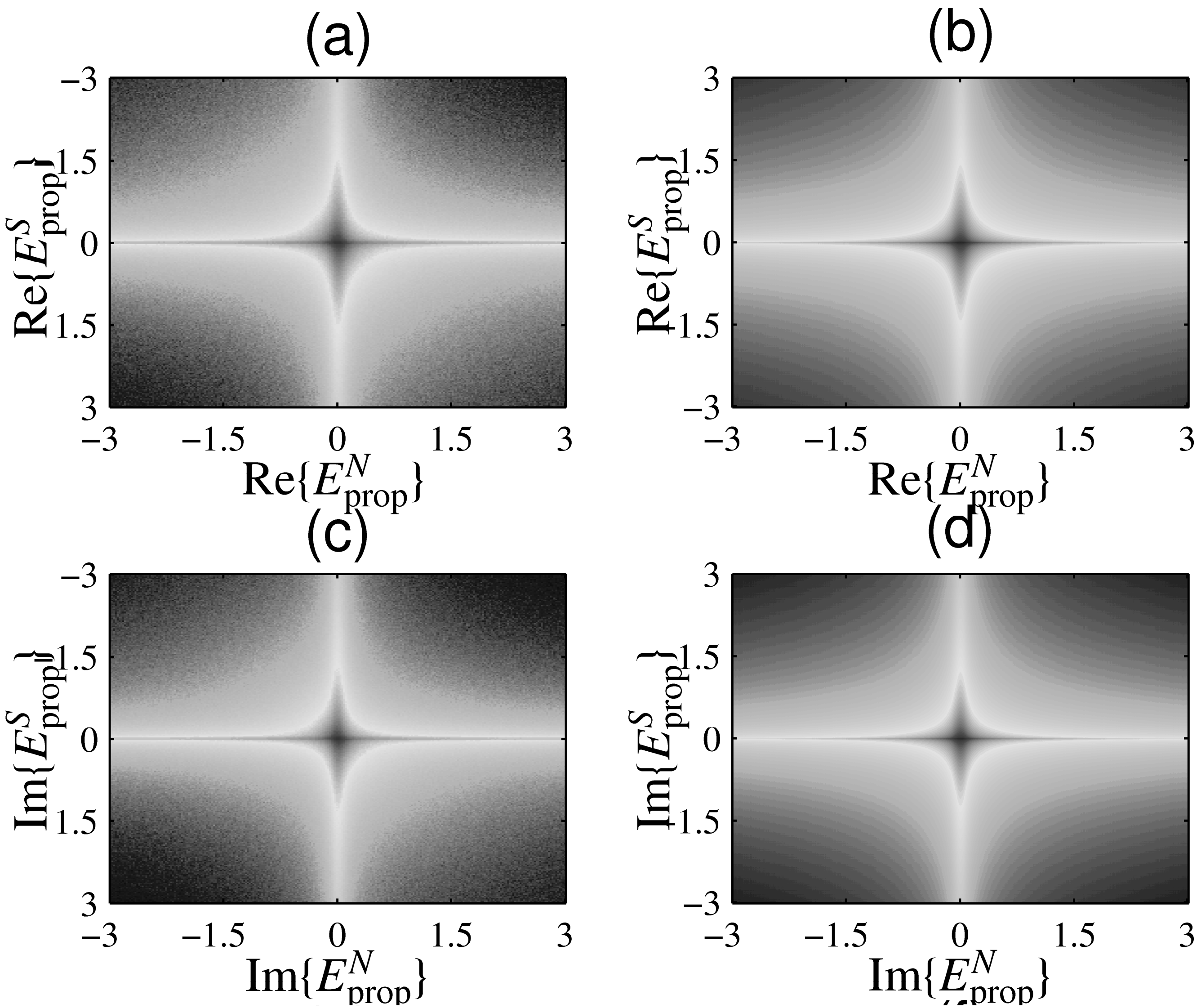
Parameter	*Settings*
Sampling frequency	8 kHz
Frame length	160 (20 ms)
FFT length	256 (including zero-padding)
Frame overlap	75% (Hann window)
Input SNR	-10 dB ... 35 dB (step size: 5 dB)

Propagation Step	
AC length L_{AC}	6
Model order N_{K}	3
Model order M_{K}	2

Update Step	
Noise estimation	Minimum Statistics
SNR estimation	Decision-directed approach

Independence Assumption of Prediction Errors

Joint distributions | Product of marginal distr.



Independence Assumption of Prediction Errors

$$D(\lambda,\mu) = E_{\mathsf{S}}(\lambda,\mu) + E_{\mathsf{N}}(\lambda,\mu)$$

- Assumption made in update step:

$$\mathbb{E}\{|D(\lambda,\mu)|^2\} = \mathbb{E}\{|E_{\mathsf{S}}(\lambda,\mu)|^2\} + \mathbb{E}\{|E_{\mathsf{N}}(\lambda,\mu)|^2\}$$

- Introduced error using this assumption

$$\mathsf{LogERR} = 10 \cdot \log_{10}\left(\frac{\mathbb{E}\{|D(\lambda,\mu)|^2\}}{\mathbb{E}\{|E^S_{\mathsf{prop}}(\lambda,\mu)|^2\} + \mathbb{E}\{|E^N_{\mathsf{prop}}(\lambda,\mu)|^2\}}\right)$$

SNR	-10 dB	-5 dB	0 dB	5 dB	10 dB
$\widehat{\mathsf{LogERR}}$	0.0097 dB	0.0112 dB	0.0120 dB	0.0108 dB	0.0097 dB

SNR	15 dB	20 dB	25 dB	30 dB	35 dB
$\widehat{\mathsf{LogERR}}$	0.0072 dB	0.0052 dB	0.0034 dB	0.0027 dB	0.0022 dB

Parameter settings for complex DFT estimator

- Generalized Gamma PDF assumed for speech prediction error [Erkelens et al., IEEE SigPro Letters 08]

$$p_{|E_S|}(x) = \frac{\gamma\delta^\nu}{\Gamma(\nu)} x^{\gamma\nu-1} \exp(-\delta x^\gamma)$$

with $\delta > 0$, $\gamma > 0$, $\nu > 0$ and $0 \le x < \infty$

- Resulting parameter settings

SNR [dB]	≤ -20	-15	-10	-5	0	5
γ	1	1	1	1	1	1
ν	1.41	1.05	0.87	0.76	0.72	0.67

SNR [dB]	10	15	20	25	30	≥ 35
γ	1	1	1	1	1	1
ν	0.63	0.60	0.57	0.54	0.52	0.50

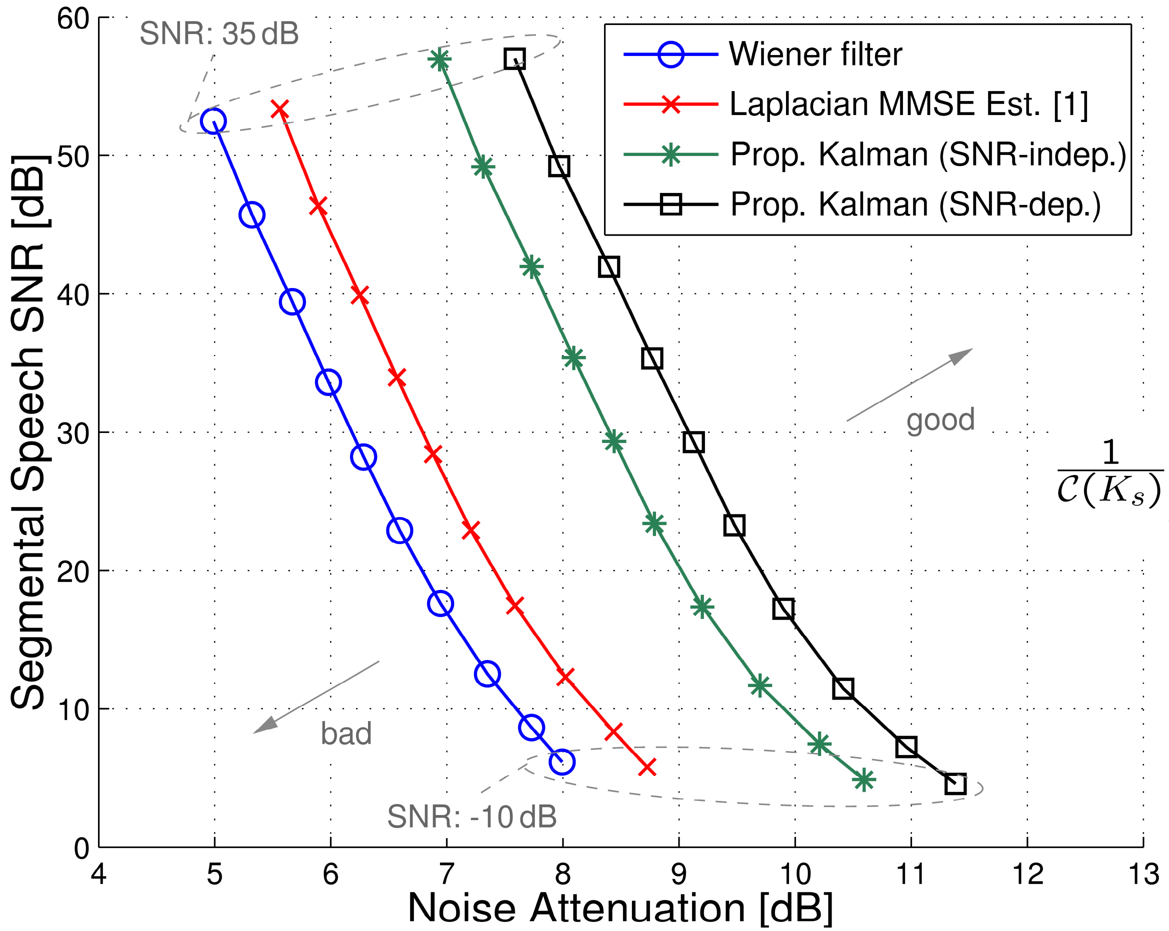
Evaluation, Test Results, and Demonstration

5 dB f16 noise	10 dB factory noise	
		Noisy
		Wiener Filter
		Laplacian MMSE Estimator [1]
		Modified Kalman Filter (SNR-independent)
		Modified Kalman Filter (SNR-dependent)

[1] Martin and Breithaupt, IWAENC 03

Evaluation, Test Results, and Demonstration

- Objective measurements (f16, babble, car, factory, WGN)



Noise attenuation:

$$10 \cdot \log\left(\frac{1}{\mathcal{C}(K_n)} \sum_{k \in K_n} \frac{\mathbb{E}\{n^2(k)\}}{\mathbb{E}\{\tilde{n}^2(k)\}}\right)$$

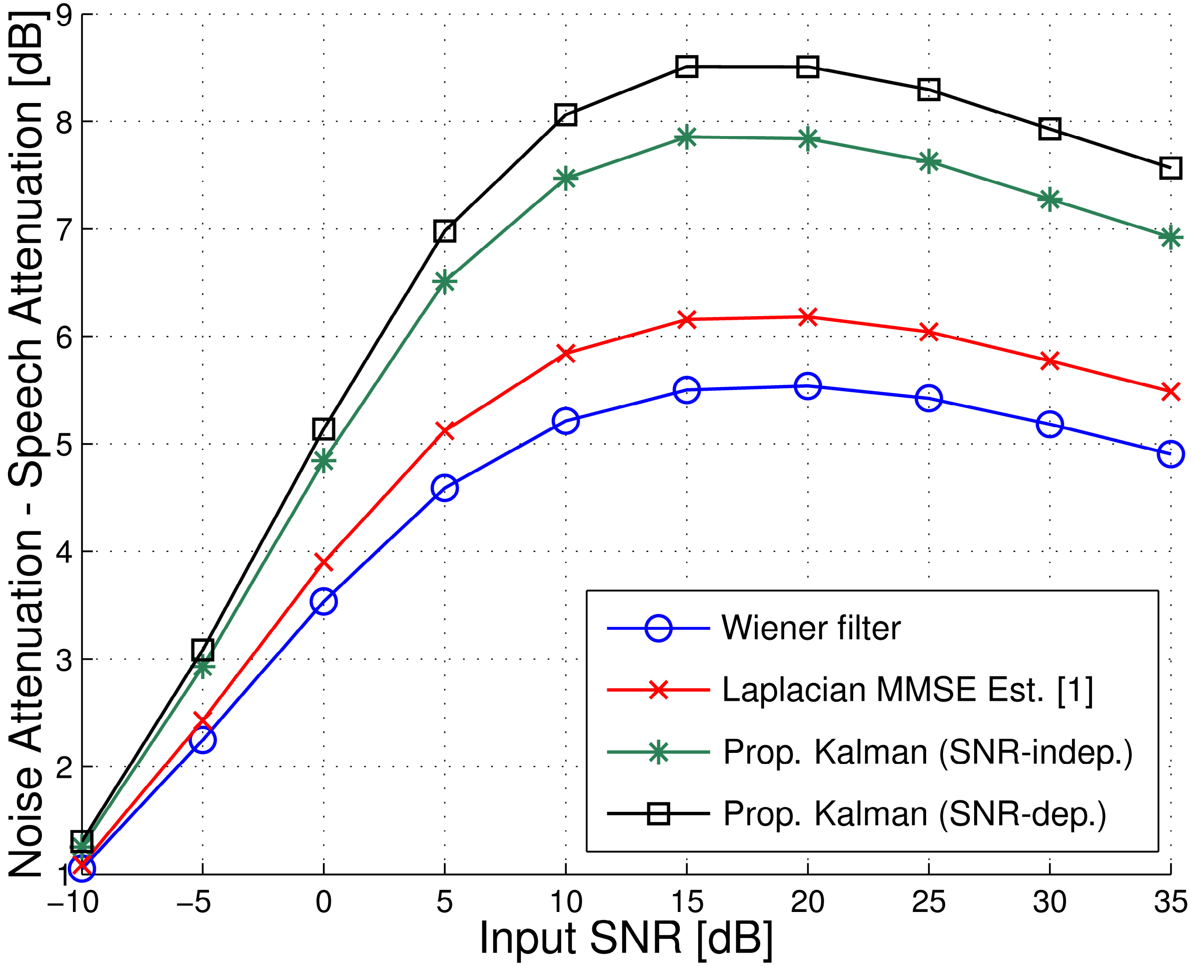
Seg. speech SNR:

$$\frac{1}{\mathcal{C}(K_s)} \sum_{\lambda \in K_s} 10 \cdot \log\left(\frac{\sum_{\nu=0}^{M-1} s^2(\nu+\lambda M)}{\sum_{\nu=0}^{M-1}\left(\tilde{s}(\nu+\lambda M)-s(\nu+\lambda M)\right)^2}\right)$$

$\tilde{s}(k)$	filtered speech signal
$\tilde{n}(k)$	filtered noise signal
K_s	set corresp. to speech activity
K_n	set corresp. to noise activity
$\mathcal{C}(\cdot)$	number of set elements
M	Frame length

Evaluation, Test Results, and Demonstration

- Objective measurements (f16, babble, car, factory, WGN)



[1] Martin and Breithaupt, IWAENC 03