11. ITG Fachtagung Sprachkommunikation

Spectral Noise Tracking for Improved Nonstationary Noise Robust ASR

Aleksej Chinaev, Marc Puels, Reinhold Haeb-Umbach

Department of Communications Engineering
University of Paderborn, Germany

September 24th, 2014

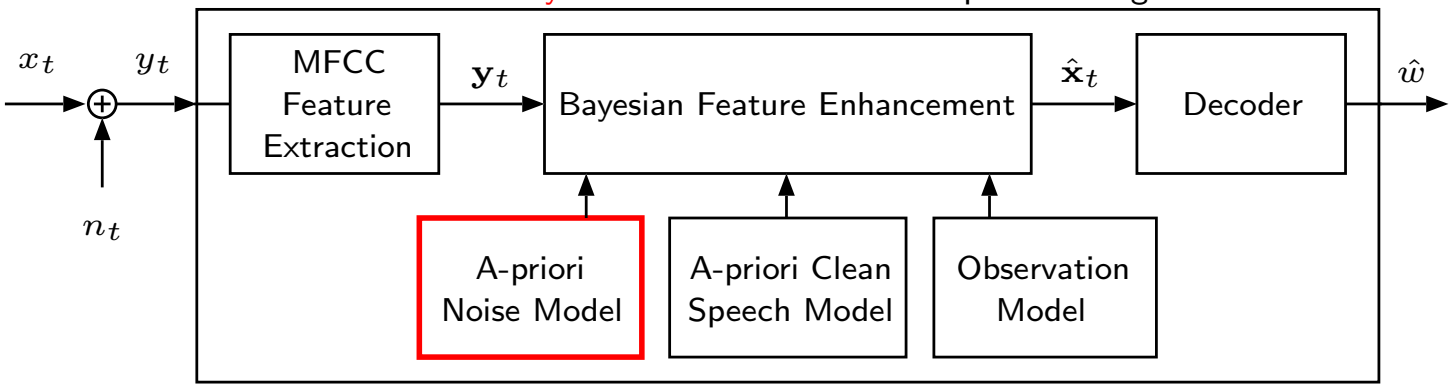
Table of Contents

Introduction

Noise Robust ASR

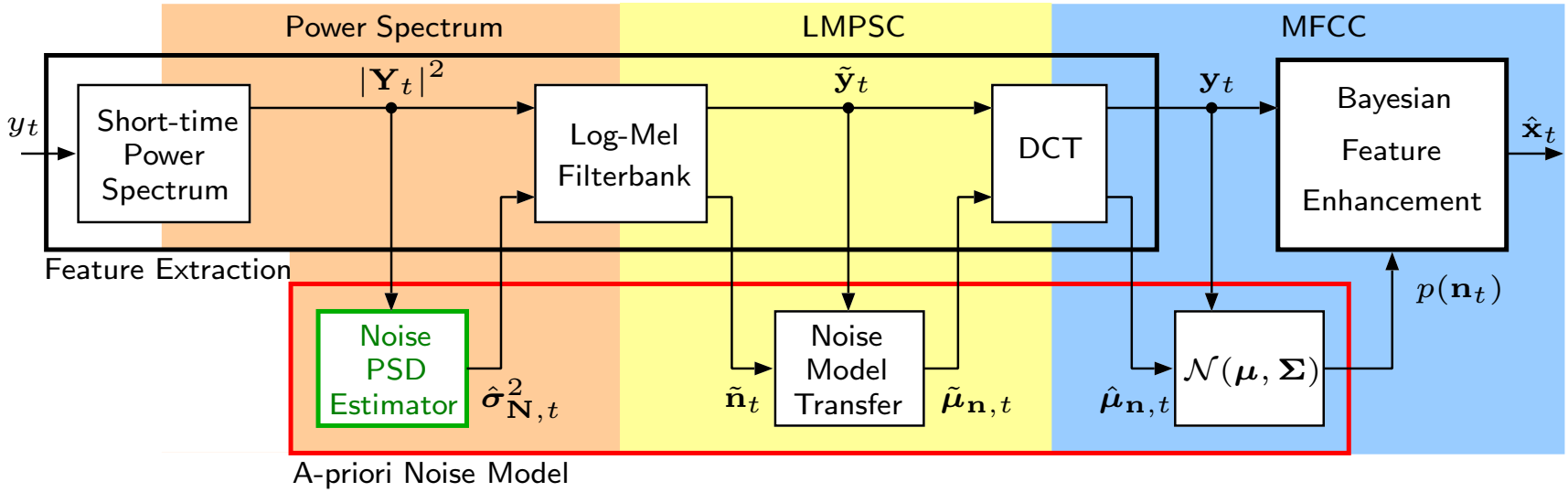
- Type of distortion - an additive nonstationary noise n_t from Aurora IV database
- ASR method - a causal Bayesian Feature Enhancement, [Leutnant et al., 2011]

Nonstationary Noise Robust Automatic Speech Recognition



- Until now: a time-invariant a-priori noise model $p(\mathbf{n}_t) = \mathcal{N}(\mathbf{n}_t; \boldsymbol{\mu}_\mathbf{n}, \boldsymbol{\Sigma}_\mathbf{n})$
  - Estimate $\boldsymbol{\mu}_\mathbf{n}, \boldsymbol{\Sigma}_\mathbf{n}$ on the speech-free frames in the beginning of an utterance
- New: $p(\mathbf{n}_t) = \mathcal{N}(\mathbf{n}_t; \boldsymbol{\mu}_{\mathbf{n},t}, \boldsymbol{\Sigma}_\mathbf{n})$ with a time-variant mean vector $\boldsymbol{\mu}_{\mathbf{n},t}$

A-priori model of nonstationary noise for a Bayesian framework



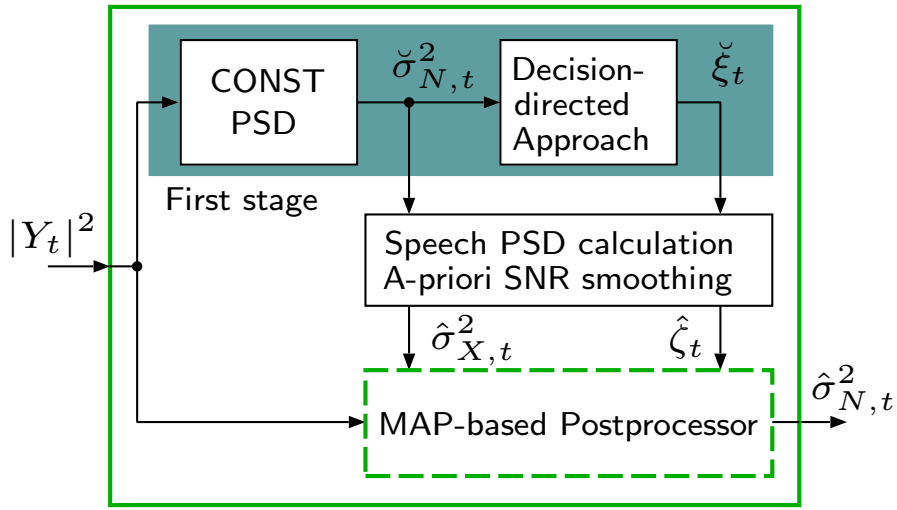
Stepwise calculation of the a-priori noise model $p(\mathbf{n}_t) = \mathcal{N}(\mathbf{n}_t; \boldsymbol{\mu}_{\mathbf{n},t}, \boldsymbol{\Sigma}_{\mathbf{n}})$

1. Estimate a time-variant power spectral density (PSD) $\boldsymbol{\sigma}^2_{\mathbf{N},t} = E[|\mathbf{N}_t|^2]$ of the nonstationary noise signal from a noisy power spectrum $|\mathbf{Y}_t|^2$
2. Calculate a time-variant mean vector $\tilde{\boldsymbol{\mu}}_{\mathbf{n},t}$ from estimates $\tilde{\mathbf{n}}_t$ in the LMPSC domain by using the Noise Model Transfer approach
3. Estimate the time-invariant covariance matrix $\boldsymbol{\Sigma}_{\mathbf{n}}$ for the Gaussian noise model

Maximum a-posteriori (MAP) based spectral noise tracking

Signal processing in the first stage

- Time-invariant noise PSD estimator (CONST PSD) based on the speech-free frames in the beginning of an utterance
- Decision-directed approach for estimation of the a-priori SNR $\breve{\xi}_t$, [Ephraim et al., 1984]



Noise PSD Estimator for one frequency bin

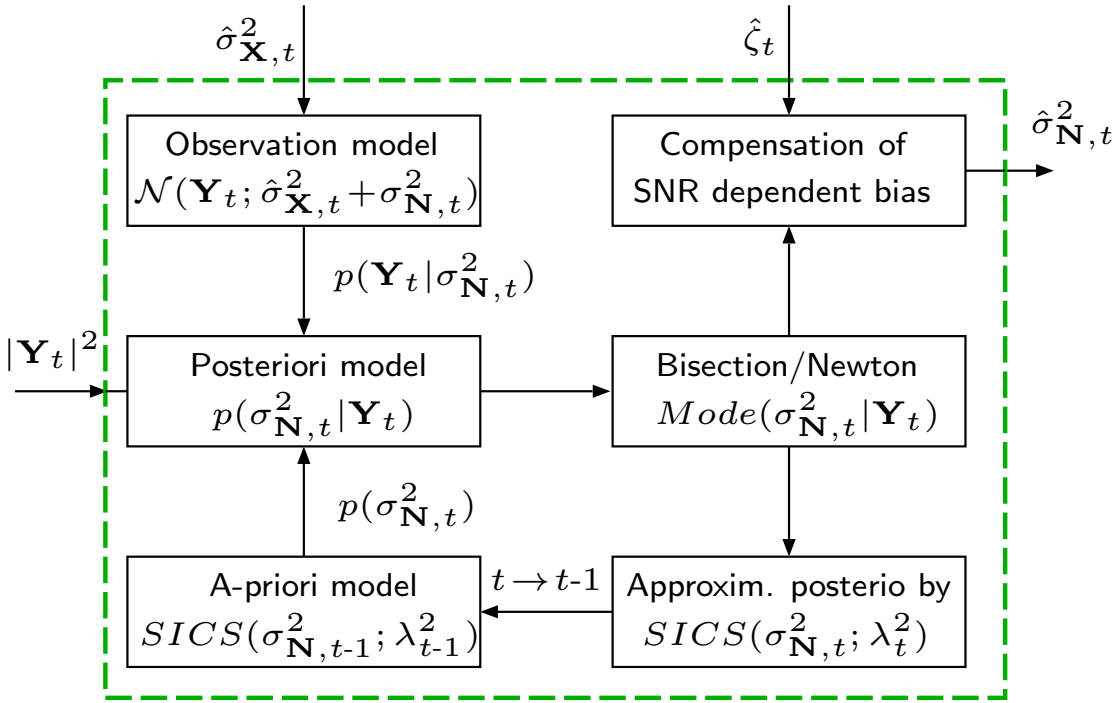
MAP-based (MAP-B) noise PSD tracker as a postprocessor

- Given the noisy power $|Y_t|^2$, the clean speech PSD $\hat{\sigma}^2_{X,t}$ and the a-priori SNR $\hat{\zeta}_t$ calculate a MAP-B noise PSD estimate $\hat{\sigma}^2_{N,t}$, [Chinaev et al., 2012]

Maximum A-Posteriori Based (MAP-B) noise PSD tracker

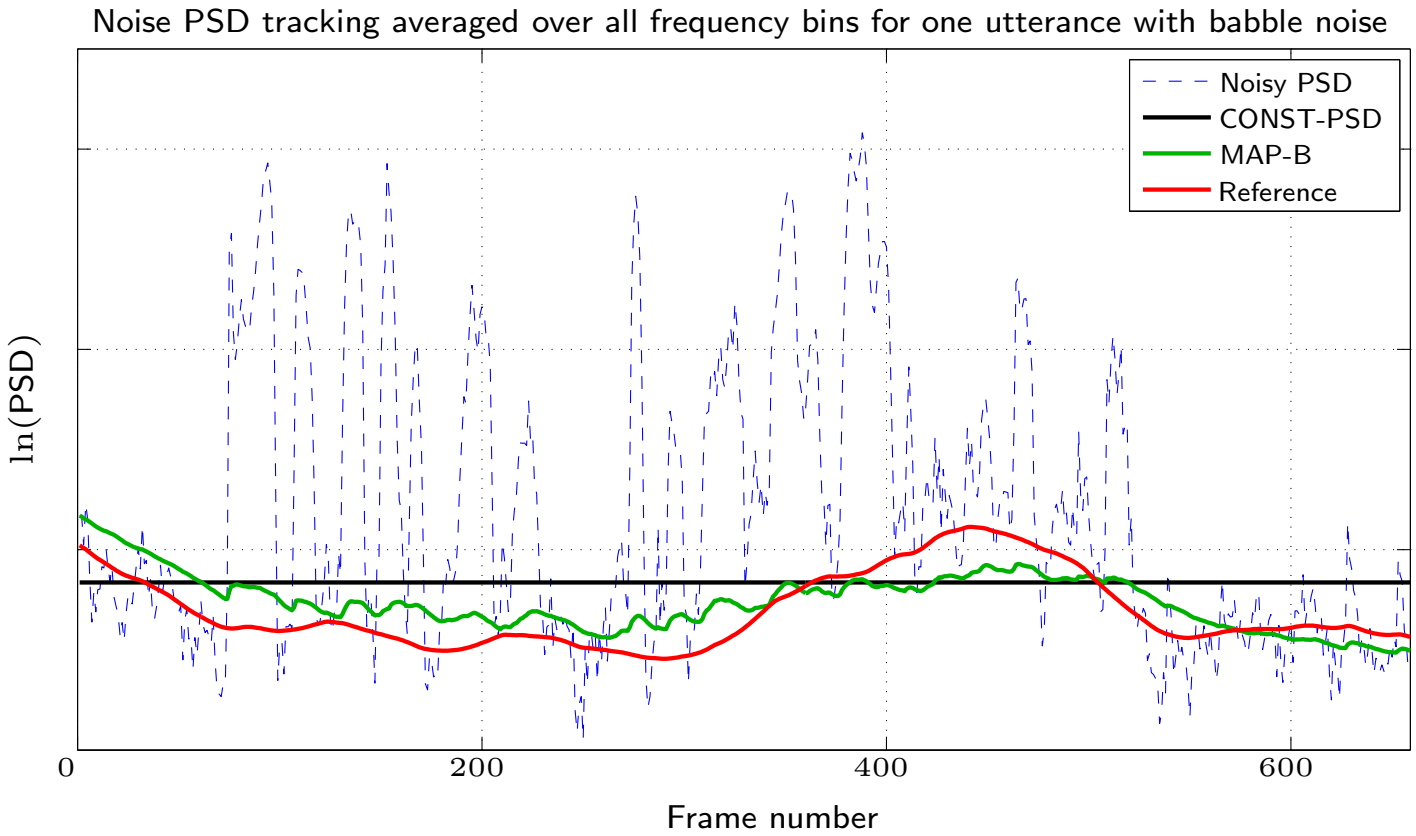
Noise PSD estimation even in presence of speech signal

- For calculation of the MAP-B estimate $\hat{\sigma}^2_{\mathbf{N},t}$ model observations $\mathbf{Y}_t$ by the Gaussian distribution $p(\mathbf{Y}_t|\sigma^2_{\mathbf{N},t})$ and the noise PSD $\sigma^2_{\mathbf{N},t}$ by the scaled inverse chi-squared (SICS) distribution $p(\sigma^2_{\mathbf{N},t})$



MAP-B noise PSD tracker is a core component of the Noise PSD Estimator for one frequency bin

An example for improved noise tracking in power spectral domain

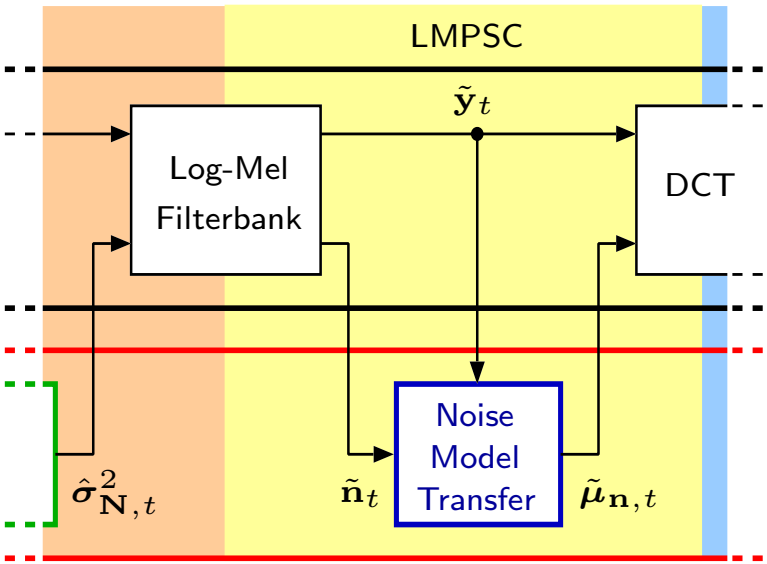


- Noise PSD estimates of the MAP-B postprocessor aim to follow the Reference

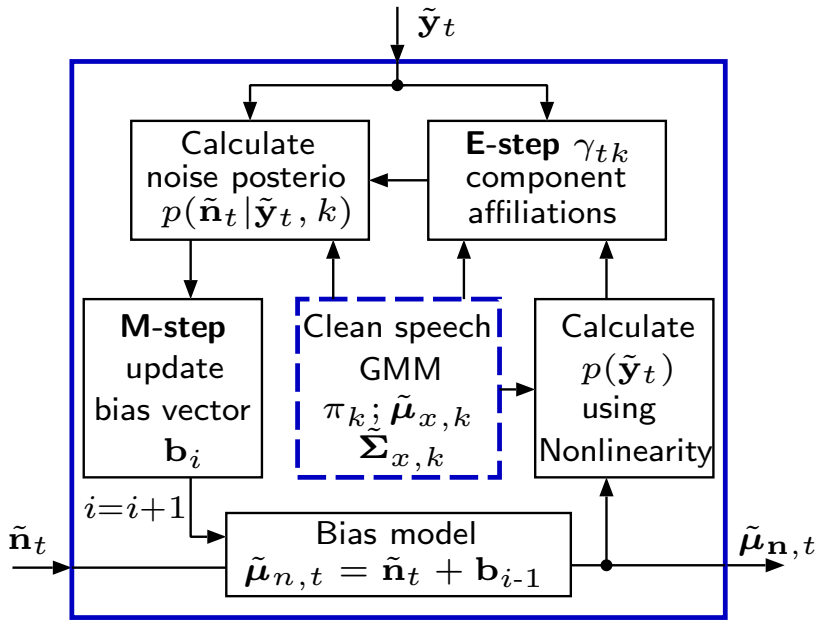
Noise Model Transfer (NMT) approach

Correction of the estimates $\tilde{\mathbf{n}}_t$ in the LMPSC domain

- Assumed a bias model $\tilde{\boldsymbol{\mu}}_{n,t} = \tilde{\mathbf{n}}_t + \mathbf{b}$ estimate a time-invariant bias vector $\mathbf{b}$ for each utterance by using the EM approach, [Yoshioka et al., 2013]



A-priori Noise Model



One EM iteration of the Noise Model Transfer

Noise tracking in the LMPSC domain on the Aurora IV database

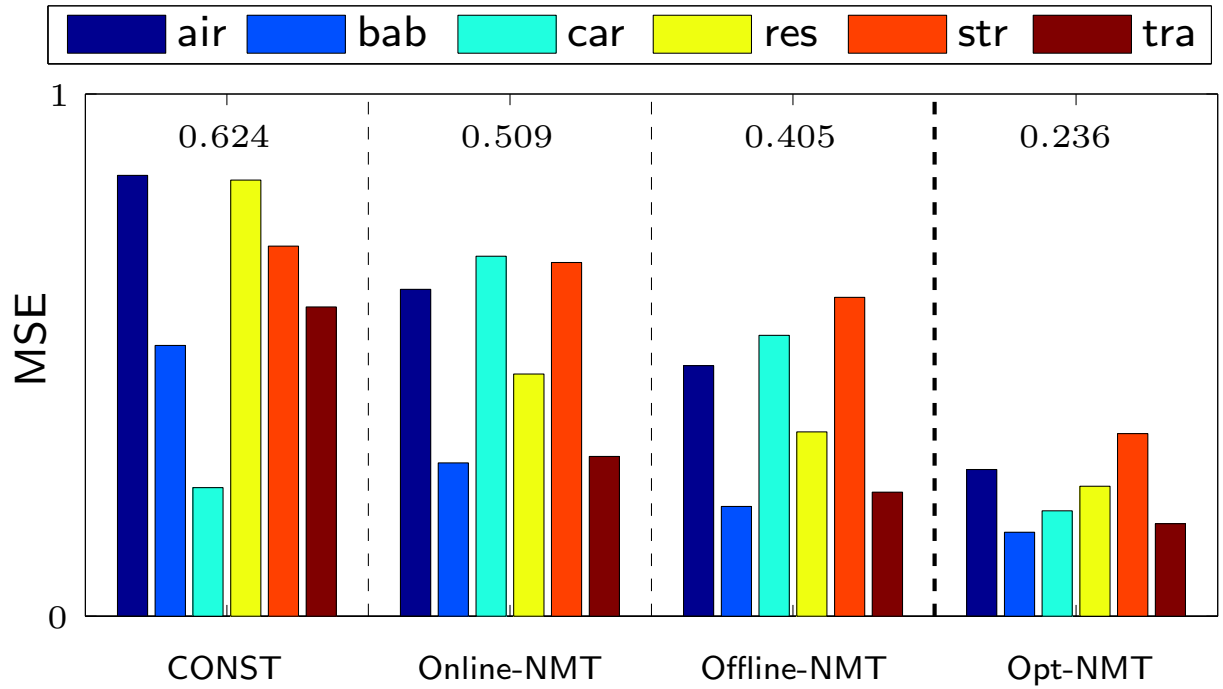


Figure: Averaged mean squared error (MSE) values of different noise LMPSC estimates $\tilde{\boldsymbol{\mu}}_{\mathbf{n},t}$

- Online-NMT: vector $\mathbf{b}$ is calculated based on the previous data $\tilde{\mathbf{y}}_t$ for $t \in [1; t]$
- Offline-NMT: conventional NMT approach based on data of the hole utterance
- Opt-NMT: using the true bias vector $\mathbf{b}_{\text{true}}$

Recognition results on the Aurora IV database

	Baseline	CONST	Online-NMT	Offline-NMT	Opt-NMT	Oracel
clean	12.7	13.0	12.0	12.1	12.9	12.2
airport	61.5	51.9	47.3	47.4	44.5	29.3
babble	60.6	47.0	42.6	43.0	42.0	32.0
car	39.0	19.5	17.1	16.9	17.6	15.9
restaurant	58.8	52.7	51.9	50.5	46.9	34.8
street	58.2	43.5	43.9	42.1	41.4	30.9
train	60.6	43.0	43.0	42.6	42.0	33.7
AVG	**50.2**	**38.7**	**36.8**	**36.4**	**35.3**	**27.0**

Table: Resulting word error rates on the Aurora IV database

- Improved nonstationary noise tracking leads to a consistent decrease of the averaged (AVG) word error rates

Summary and outlook

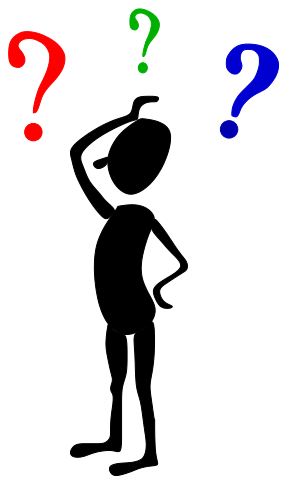
Summary

- A-priori model for nonstationary noise
  - Spectral noise tracking by using the MAP-B postprocessor
  - Transformation of the noise PSD estimates into the MFCC domain by using the Noise Model Transfer approach
  - Bayesian feature enhancement with the time-variant a-priori noise model
- Experimental results on the Aurora IV database
  - Improved tracking of nonstationary noise features
  - Consistent decrease of word error rates $\Rightarrow$ nonstationary noise robust ASR

Outlook

- Better start values for the EM approach by considering the mismatch function
- Additional usage of a time-variant covariance matrix $\mathbf{\Sigma_n} \rightarrow \mathbf{\Sigma_{n},}_t$

Thank you for your attention!



Questions?

Dipl.-Ing. Aleksej Chinaev

University of Paderborn
Department of Communications Engineering

chinaev@nt.uni-paderborn.de
nt.uni-paderborn.de