PADERBORN UNIVERSITY
The University for the Information Society

Noise-Presence-Probability-Based Noise PSD Estimation by Using DNNs

12. ITG Fachtagung Sprachkommunikation

Aleksej Chinaev, Jahn Heymann, Lukas Drude, Reinhold Haeb-Umbach

Department of Communications Engineering
Paderborn University

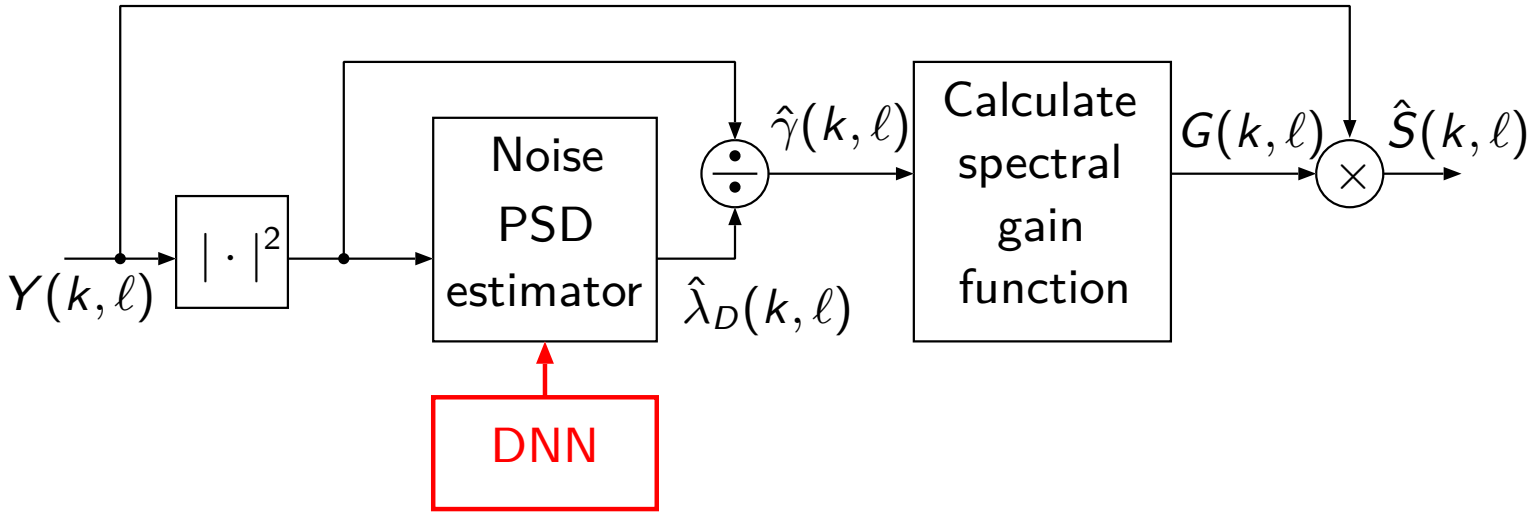
5. Oktober 2016

Table of Contents

Noise PSD estimation in spectral speech enhancement

- Clean speech $s(t)$ contaminated by an additive noise $d(t)$ resulting in

$$y(t) = s(t) + d(t) \quad \overset{STFT}{\circ\!\!-\!\!\bullet} \quad Y(k,\ell) = S(k,\ell) + D(k,\ell)$$



- Noise power spectral density (PSD) $\lambda_D(k,\ell) = E\left[|D(k,\ell)|^2\right]$ is used in a priori SNR estimation and calculation of gain function $G(k,\ell)$
- From $|Y(k,\ell)|^2$ in presence of **non-stationary** noise – challenging task

Use a Deep Neural Network (DNN) for noise PSD estimation! But how?

Techniques of 10 state-of-the-art noise PSD trackers

All estimators are causal!

noise PSD trackers	techniques: 1. SPP estimation	2. Minimum search	3. Bias compensation	4. Bayesian inference	5. Output smoothing
1. SPP-based (Hirsch-93, Gerkmann-12)	✓				✓
2. MS-based (Martin-01, Chinaev-15)		✓	✓		
3. MCRA-based (Cohen-02, Fan-07, Kum-09)	✓	✓			✓
4. IMCRA (Cohen-03)	✓	✓	✓		✓
5. MMSE-SPP (Yu-09)	✓			✓	✓
6. MMSE-BM (Hendriks-10)		✓	✓	✓	✓
proposed DNN-based	✓				✓

Table of Contents

LSTM network for causal estimation of noise spectral mask

- DNN task: estimate a soft noise spectral mask $M_{\mathrm{D}}(k,\ell)$ from $|Y(k,\ell)|$ by targeting an ideal binary mask (IBM) for noise

$$\mathrm{IBM}_D(k,\ell) = \begin{cases} 1, & |D(k,\ell)| > 10 \cdot |S(k,\ell)| \\ 0, & \text{else.} \end{cases}$$

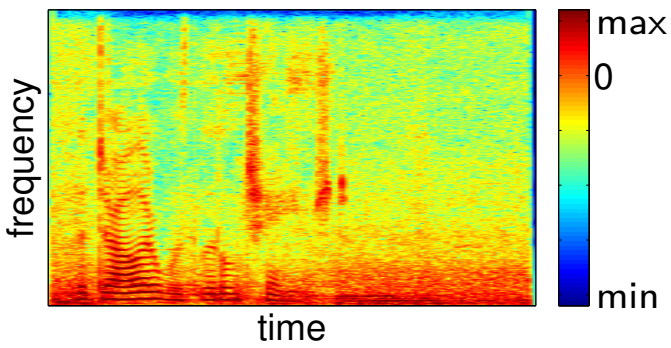
Table : Network configuration for STFT window length of 1024

Layer	Units	Type	Non-Linearity	p_{dropout}
L1	512	LSTM	Tanh	0.5
L2	1024	FF	ELU	0.5
L3	1024	FF	ELU	0.5
L4	513	FF	Sigmoid	0.0

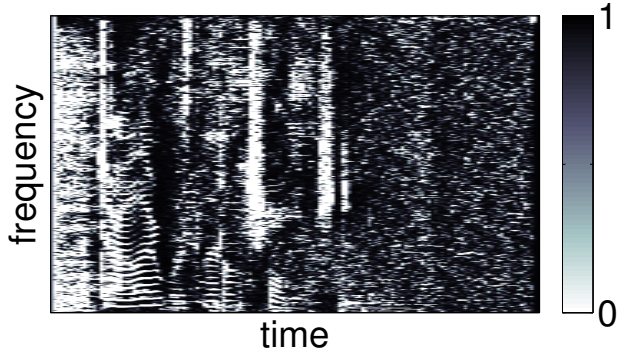
- Causality: estimation of $M_{\mathrm{D}}(k,\ell)$ based only on data of previous frames
  - $M_{\mathrm{D}}(k,\ell) \in [0,\,1]$ is soft $\rightarrow$ noise-only presence probability estimation

Example of noise-only presence probability (NPP) estimation

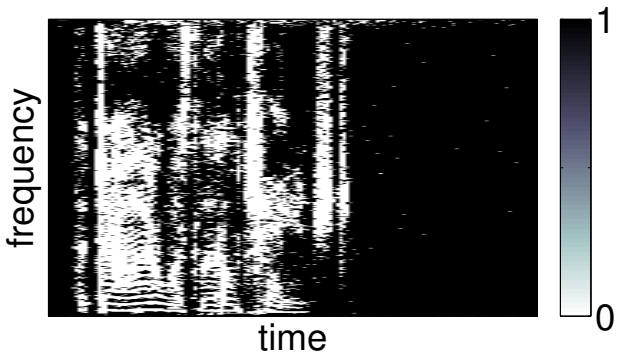
Noisy spectrogram



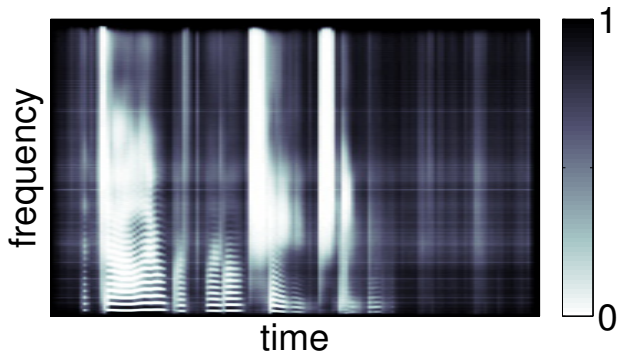
(1-SPP) von Gerkmann-12



IBM for noise



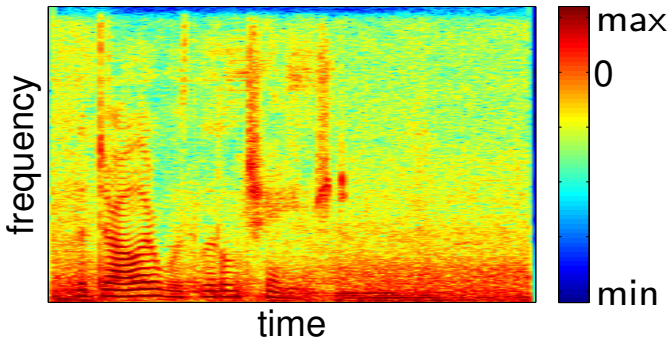
DNN-based NPP estimate



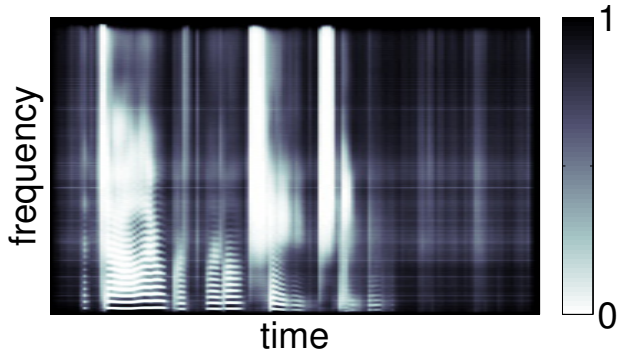
- DNN provides smoothed however structured noise spectral mask $M_{\mathrm{D}}(k, \ell)$

Output smoothing using DNN-based noise spectral mask

Noisy spectrogram



DNN-based noise spectral mask



- Low-complexity NPP-based noise PSD estimator with *Output smoothing*

$$\hat{\lambda}_D(k,\ell) = [1 - M_D(k,\ell)] \cdot \hat{\lambda}_D(k,\ell-1) + M_D(k,\ell) \cdot |Y(k,\ell)|^2$$

controlled by a soft noise spectral mask $M_D(k,\ell) \in [0,\, 1]$

- Speech presence $M_D(k,\ell) = 0$: $\hat{\lambda}_D(k,\ell) = \hat{\lambda}_D(k,\ell-1)$ $\Rightarrow$ hold
- Noise only $M_D(k,\ell) = 1$: $\hat{\lambda}_D(k,\ell) = |Y(k,\ell)|^2$ $\Rightarrow$ update

Table of Contents

Performance measures and experimental setup

Evaluation of noise PSD estimation

- Noise PSD reference: periodogram $|D(k,\ell)|^2$
- Measures: log-error mean (**LEM**) and log-error variance (**LEV**)

Impact on enhanced signal

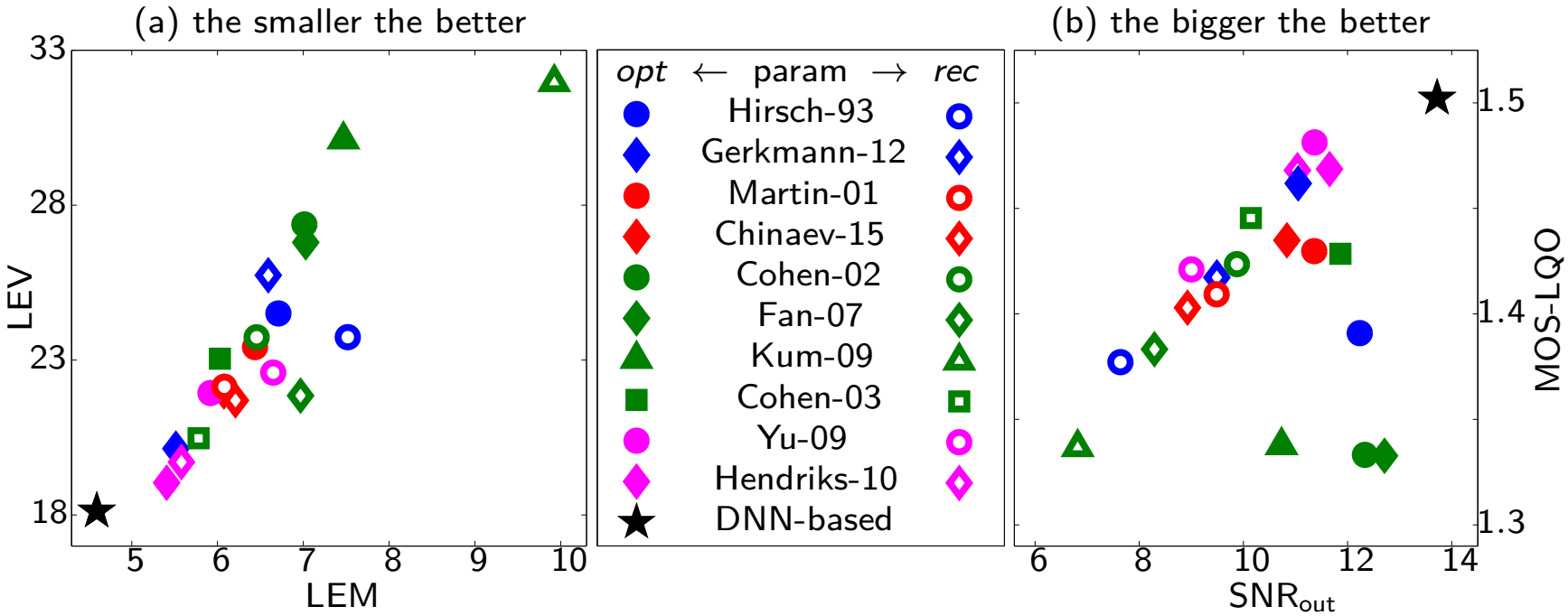
- Mean opinion score - listening quality objective (**MOS-LQO**)
- Output global signal-to-noise ratio $\mathbf{SNR}_{\text{out}}$

Development data of CHiME-3 database (5^{th} microfone)

- 4 highly non-stationary noisy environments (bus, cafe, ped, str)
- $\sim 3\,\text{h}$ of speech with averaged $\mathbf{MOS\text{-}LQO}_{\text{in}} = 1.29$ and $\mathbf{SNR}_{\text{in}} \sim 5.6\,\text{dB}$
- 25% for parameter optimization of 10 used noise PSD trackers
- 75% for comparison with the proposed DNN-based approach

Experimental results on CHiME-3 challenge

- 10 state-of-the-art noise trackers for recommended (*rec*) parameters



- Optimized (opt) parameter over all performance measures scaled on [0, 1]
  - Improvement of $\sim$ 10% in LEM and $\sim$ 24% in SNR$_{out}$
  - Length of the window for *Minimum search* $\sim$ 0.25s compared to [0.6, 1.1]s
- DNN-based approach clearly outperforms all state-of-the-art noise trackers

Table of Contents

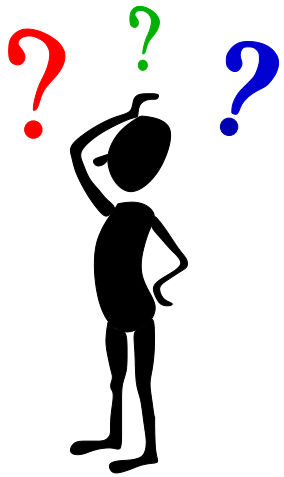
Conclusions and outlook

Conclusions

- 10 state-of-the-art noise PSD trackers are categorized and optimized
- DNN-based causal noise-only presence probability estimator proposed
- In nonstationary noise compared to considered noise PSD trackers
  - Reduced estimation error and estimator's variance
  - Better trade-off between speech quality and noise reduction

Outlook

- Applying DNN for speech presence probability estimation used in gain functions with speech presence uncertainty



Thank you for your attention!

Questions?

Paderborn University

Department of
Communications Engineering

Web: nt.upb.de