Pattern Recognition Lab

sivantos the hearing company

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF ENGINEERING

Deep Neural Networks for Noise Reduction under Hearing Aid Side Conditions

M. Aubreville[1,2], K. Ehrensperger[1], T. Rosenkranz[2], B. Graf[2], H. Puder[2], A. Maier[1]

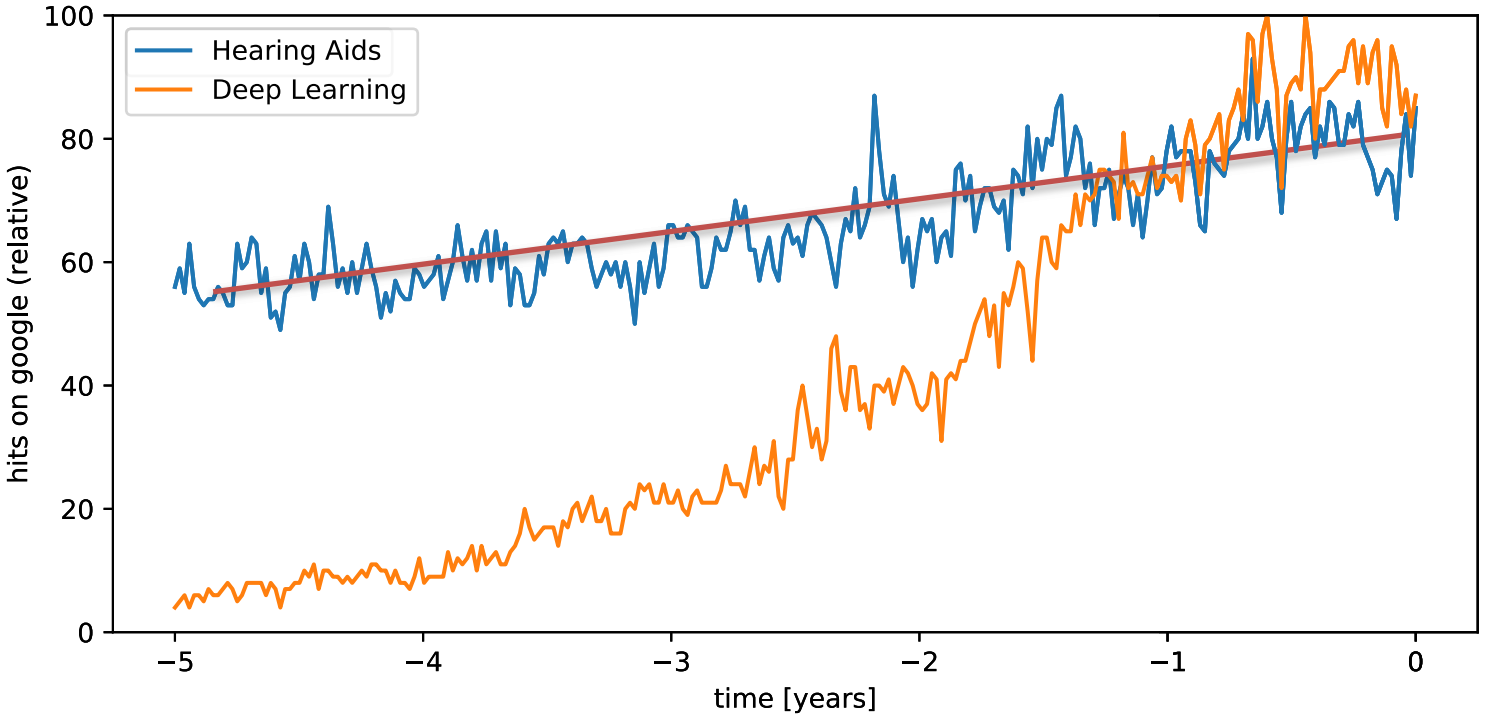
[1] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

[2] Sivantos GmbH, Erlangen, Germany

IHCON 2018, Lake Tahoe, CA, United States

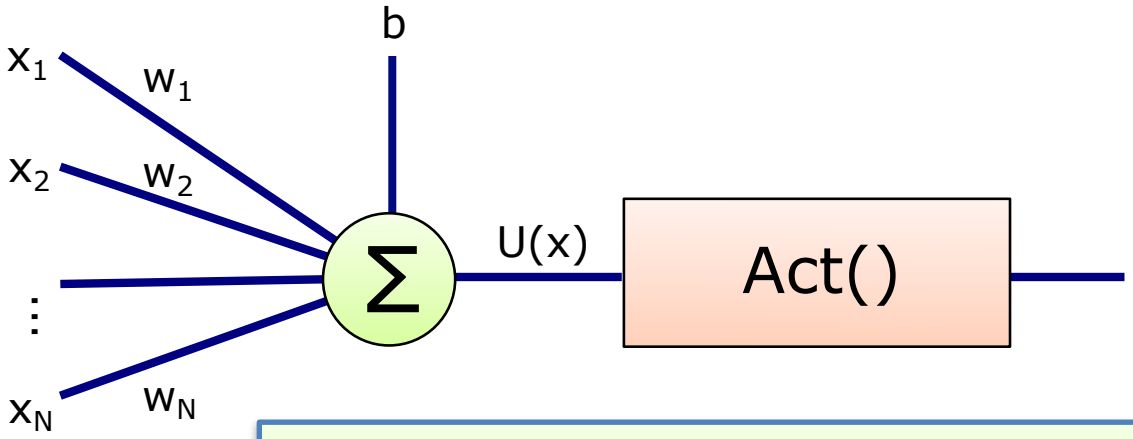
Image by the_tahoe_guy@flickr.com

Attention



[google trends: "hearing aids" vs. "deep learning"]

What is deep learning anyway?

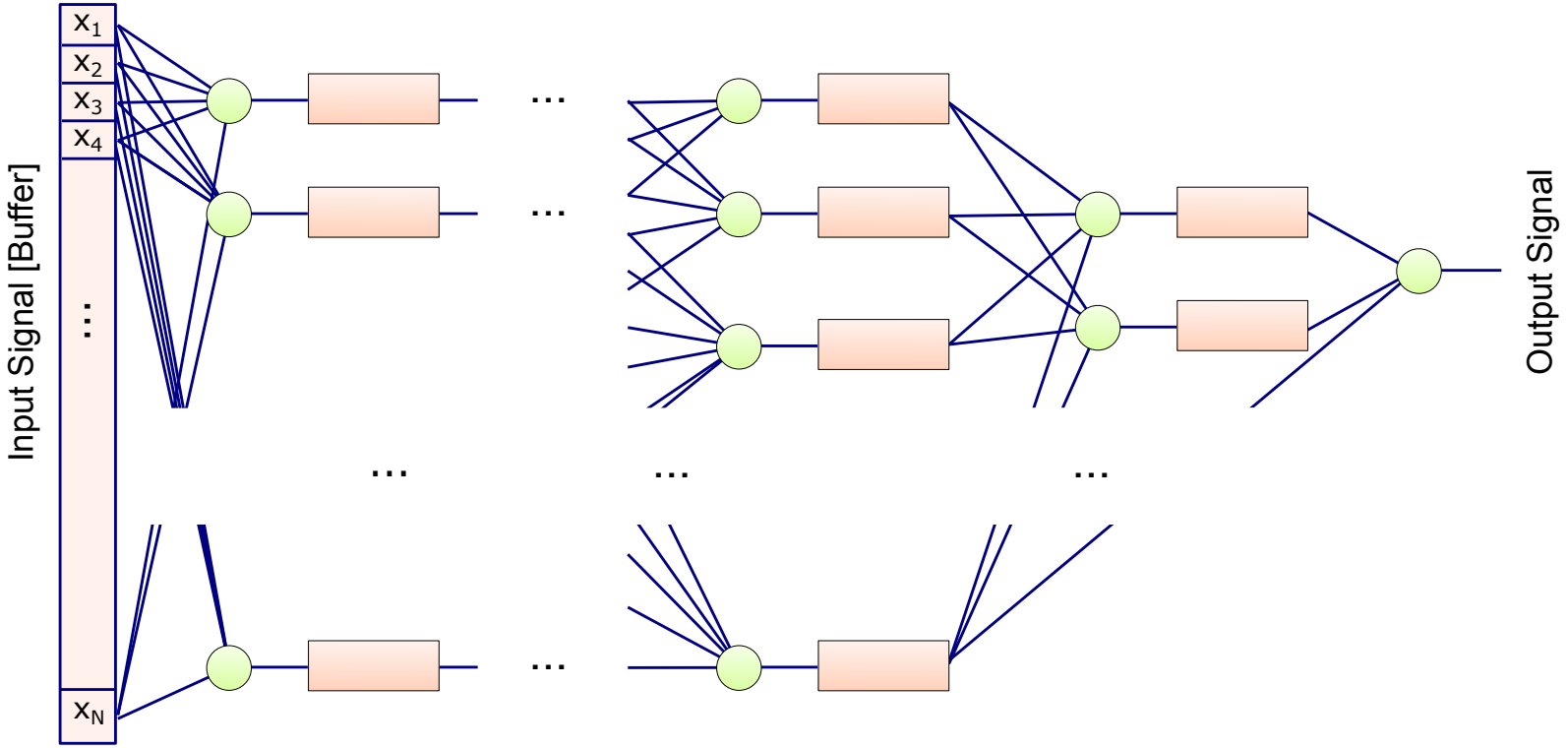


$$U(X) = b + x_1 \cdot w_1 + x_2 \cdot w_2 + \ldots + x_N \cdot w_N$$

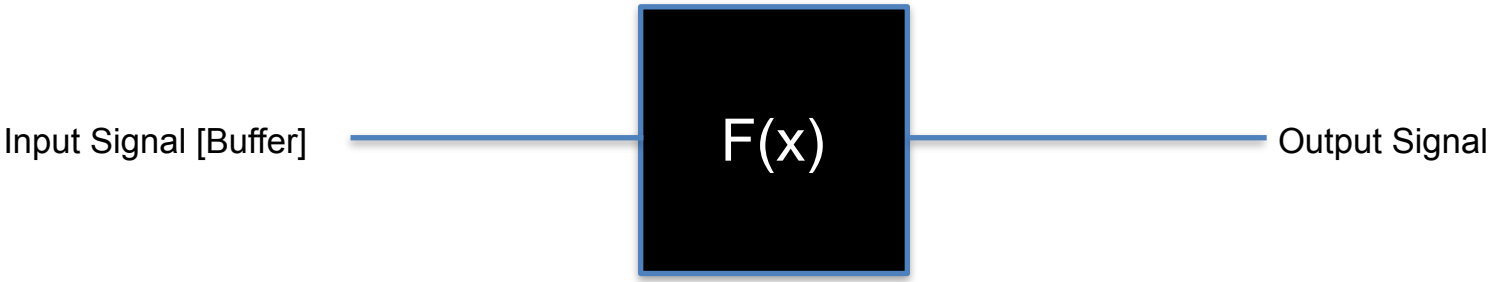
Linear Regression

Nonlinearity

What is deep learning anyway?

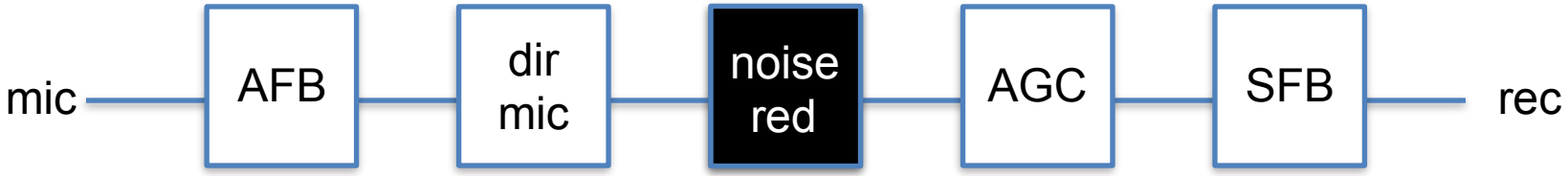


What is deep learning anyway?



- Not very smart, but very effective.
- Contents of the black box may seem unclear.
- Problem for medical devices.

Learning known Operators [MS17]



- We can exchange a single operator in the chain by a neural network.
- Restrictions to the structure of output and network help constrain the problem to be solved.

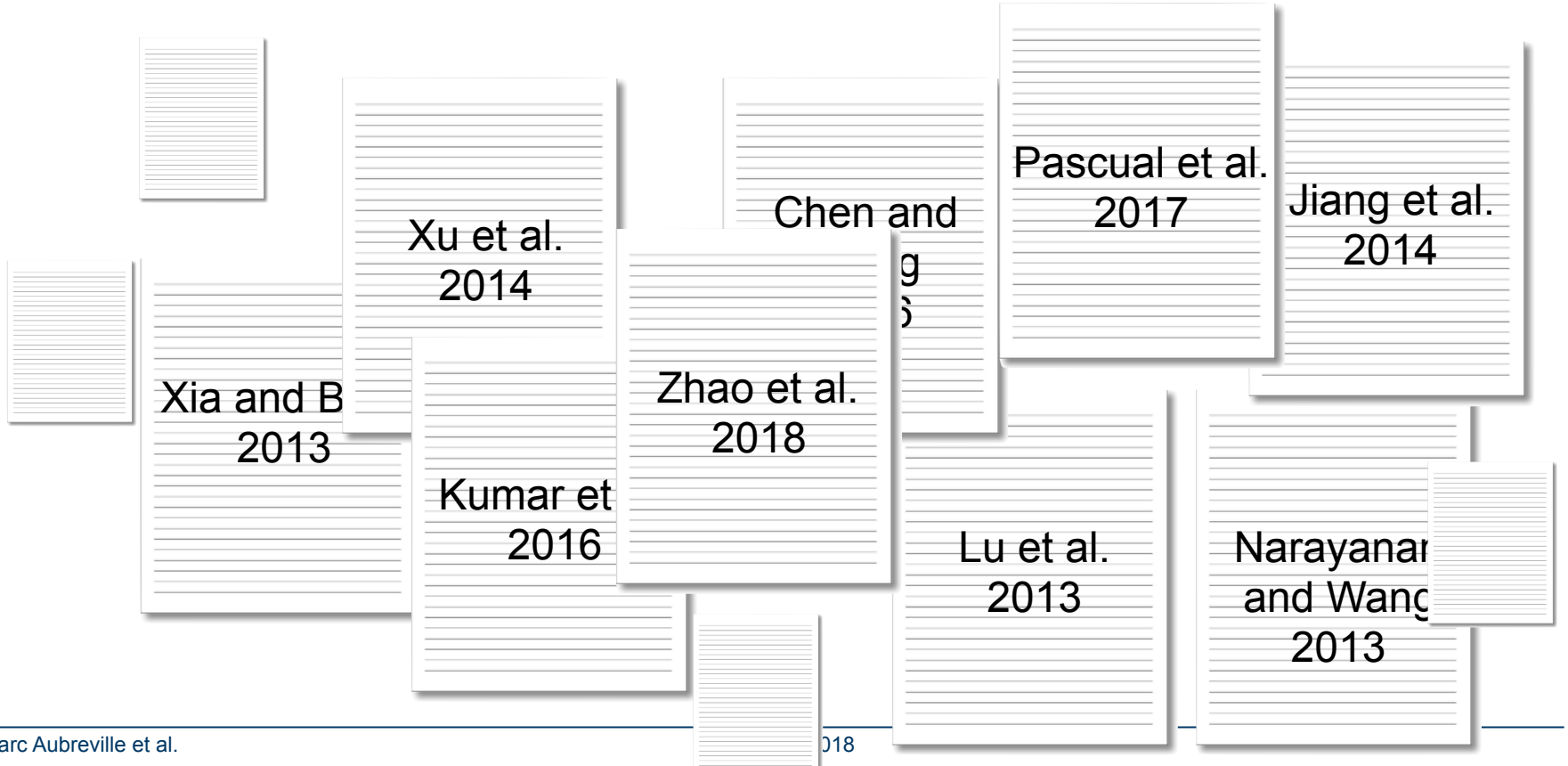
[MS17]: Maier et al., 2017: arxiv:1712.00374

Hearing Aid Side-Conditions

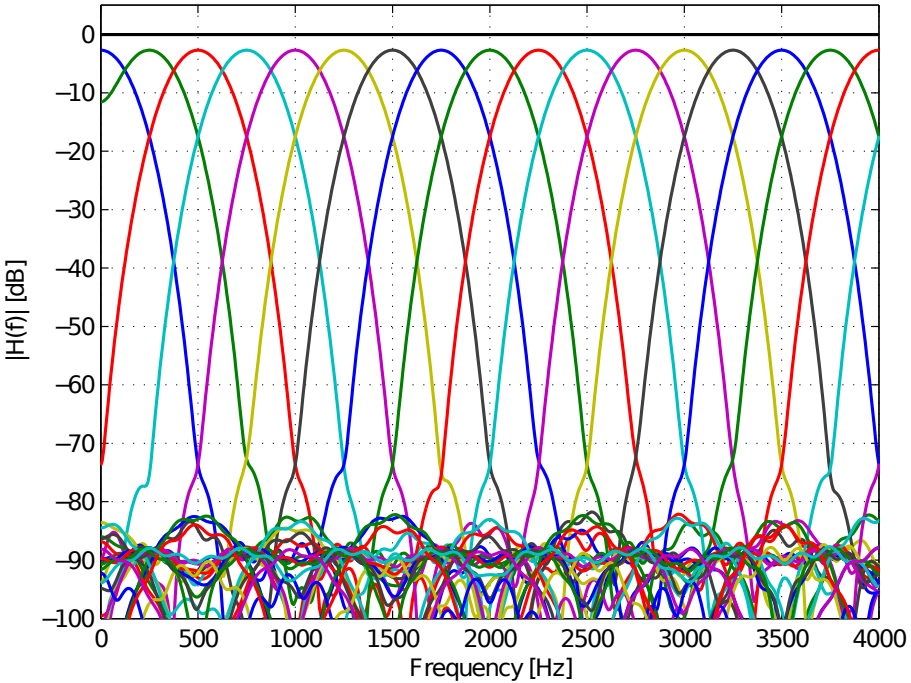
- Limited processing power
- Robustness required in any environment
- **Group delay > 10 ms objectionable to hearing aid wearers** [AT00]

[AT00]: J. Agnew and J. M. Thornton, JAAA (2000) 11:330-360

State of the (deep) art



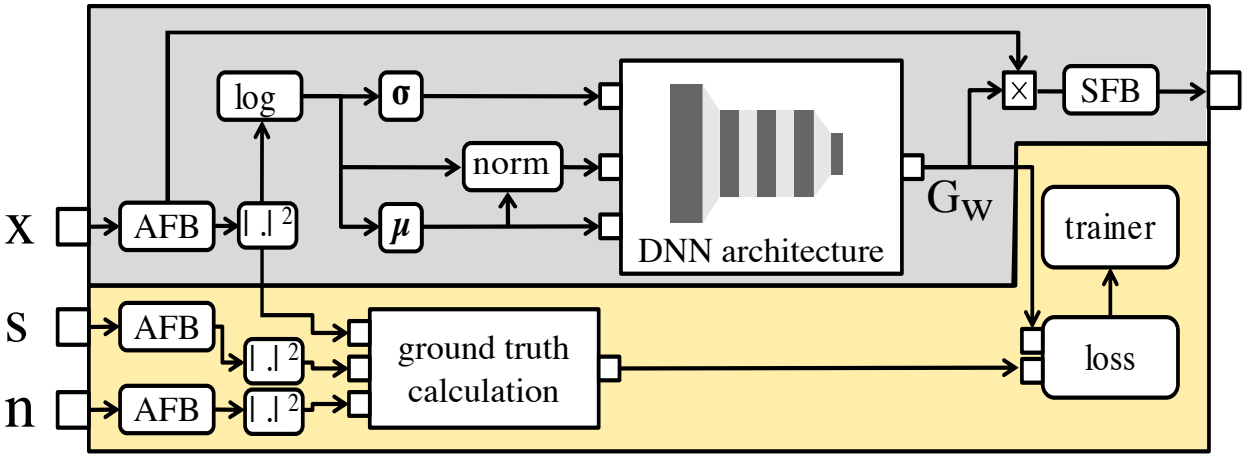
Hearing Aid Filterbank



[BS08]: R. Bäuml and W. Sörgel, EUSIPCO 2008

- State-of-the-art hearing aid filter bank
- 48 channels, uniform
- group delay: ~ 6 ms

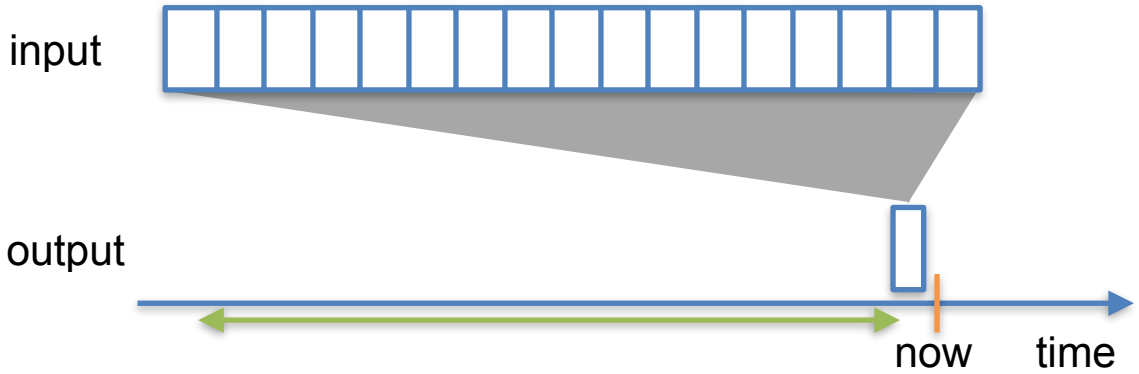
Architecture



- Input: Log-levels, normalization on window scope
- DNN with 3 hidden layers, 2048 nodes each.

Temporal context

- Temporal context should be 200-300ms, due to structure of speech (~ 4 Hz syllable rate). [HS77]
- Symmetrical context leads to high latency
- Proposal: Asymmetrical context with rich past knowledge

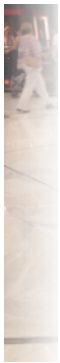


[HS77]: Houtgast and Steeneken, JASA 1985:77(3)

Our training database

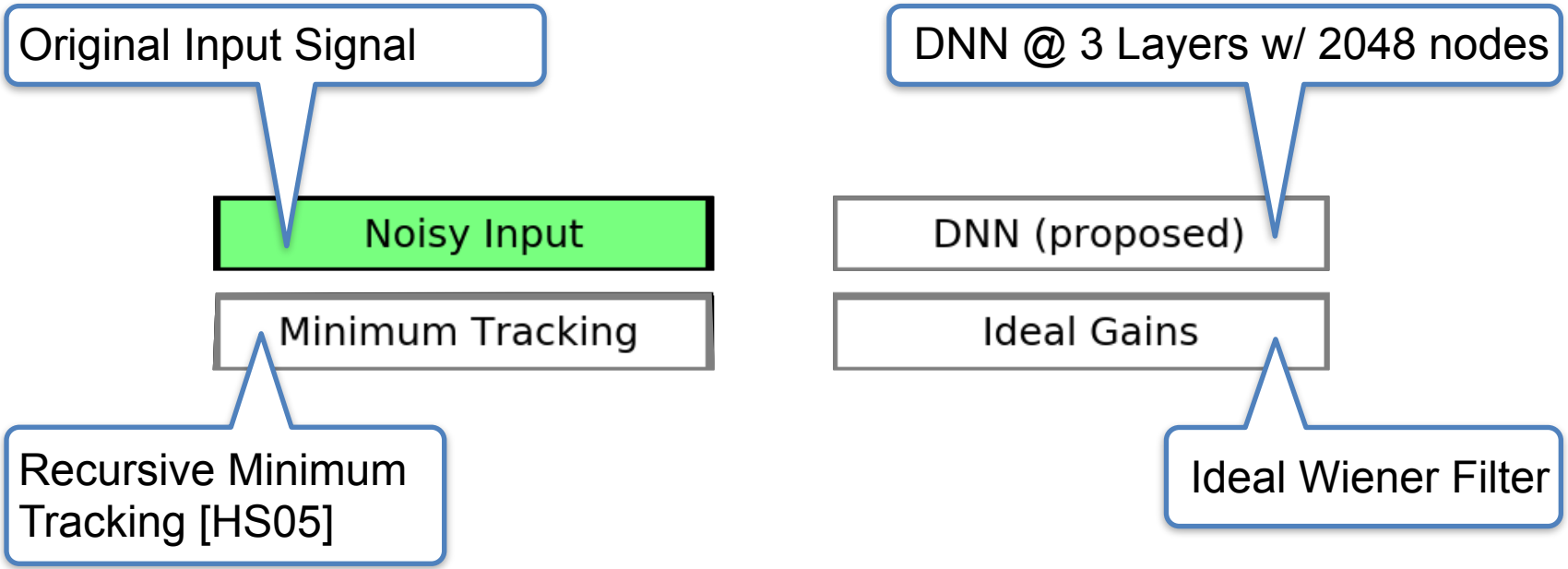
- 49 real-world noise recordings
  - recorded with hearing aid microphones
  - Mixed to achieve multi-noise conditions (Kumar et al.)



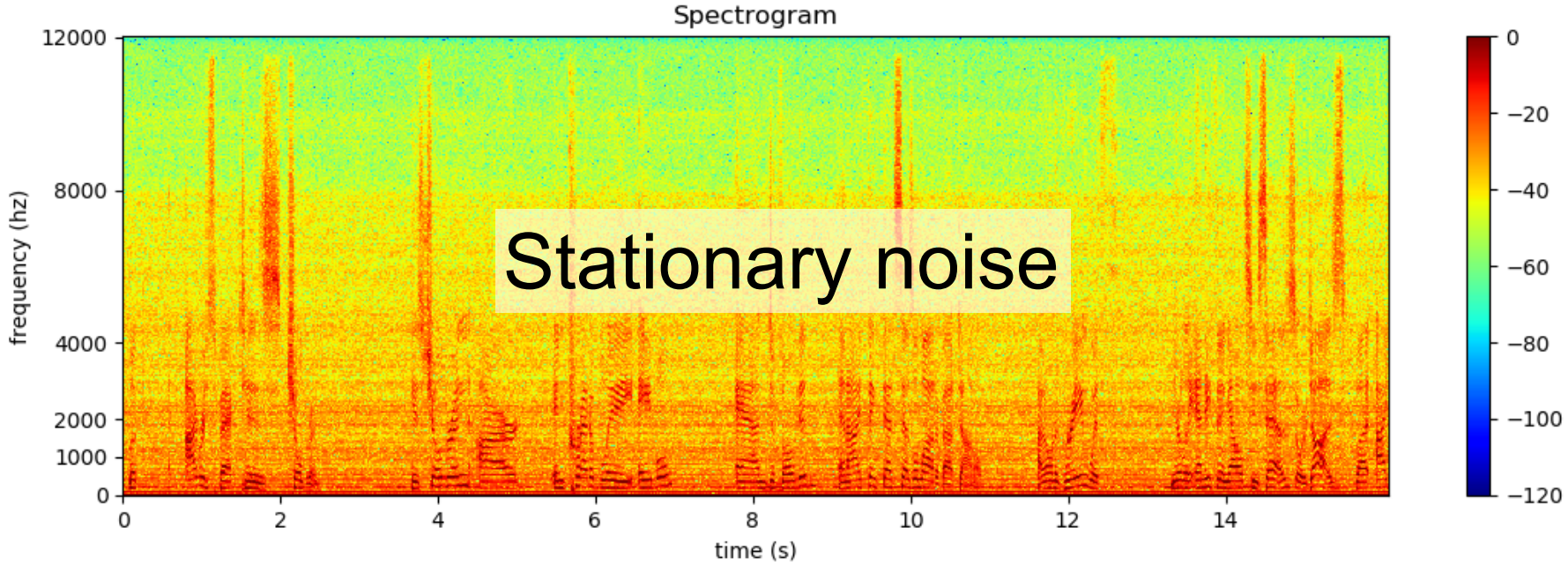


- 260 clean speech signals (EUROM, german sentences)
- Train/test split on source signal level.

Results - Conditions



[HS05]: E. Hänsler, G. Schmidt, Wiley&Sons, 2005

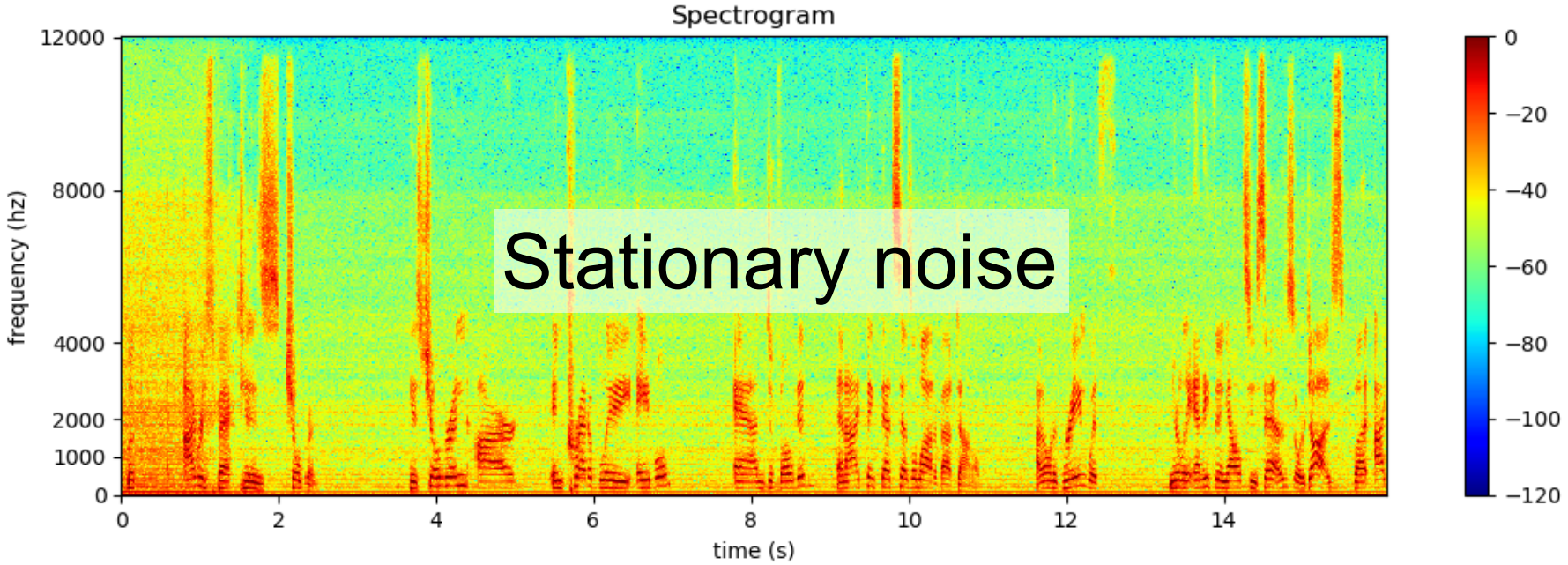


Noisy Input

DNN (proposed)

Minimum Tracking

Ideal Gains

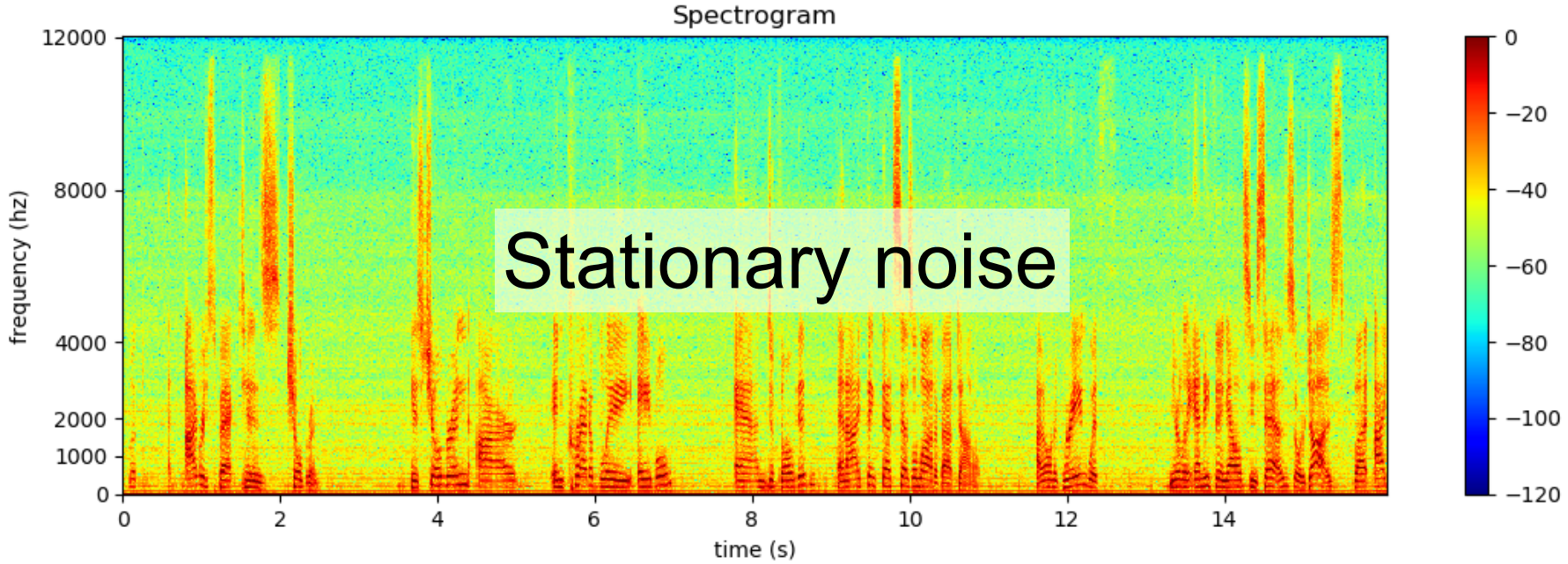


Noisy Input

DNN (proposed)

Minimum Tracking

Ideal Gains

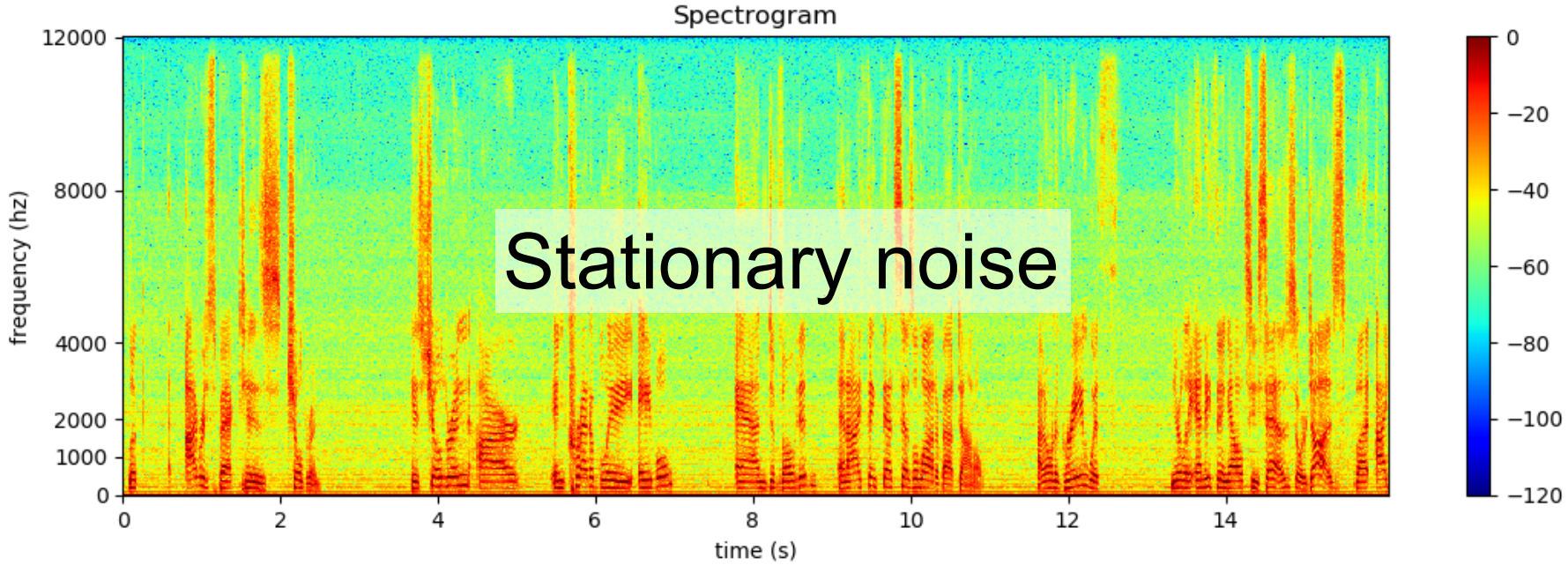


Noisy Input

DNN (proposed)

Minimum Tracking

Ideal Gains

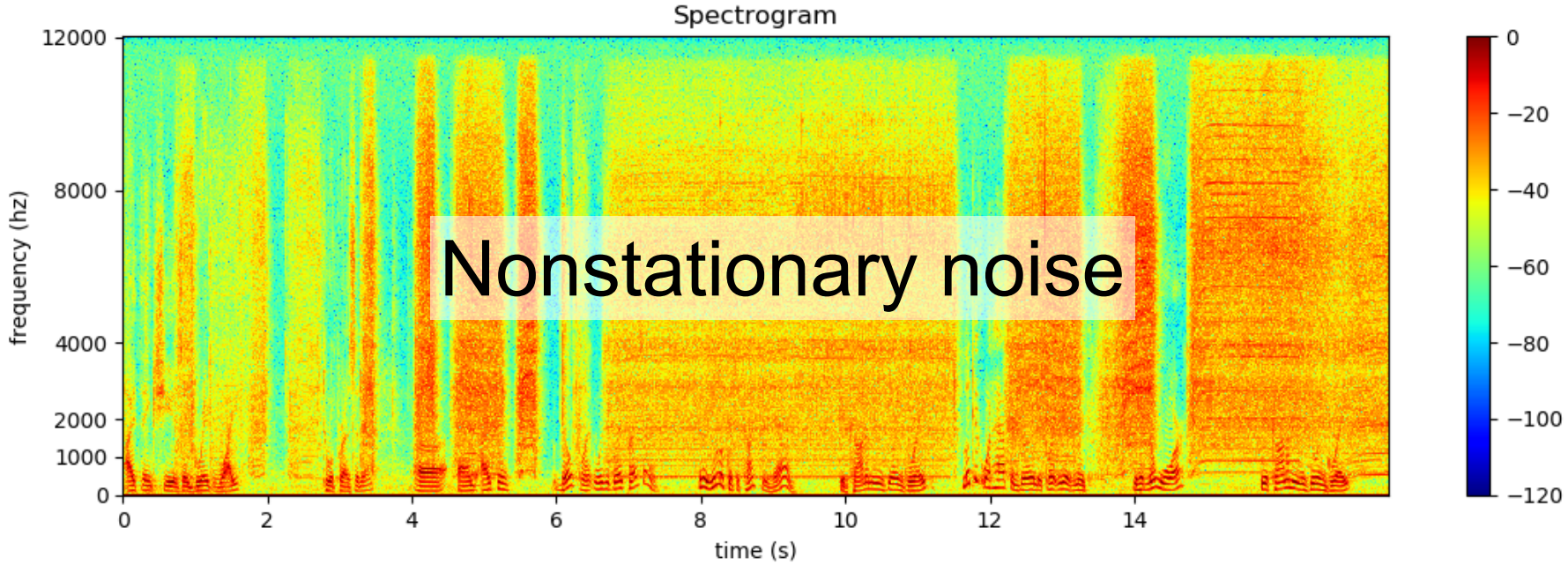


Noisy Input

DNN (proposed)

Minimum Tracking

Ideal Gains

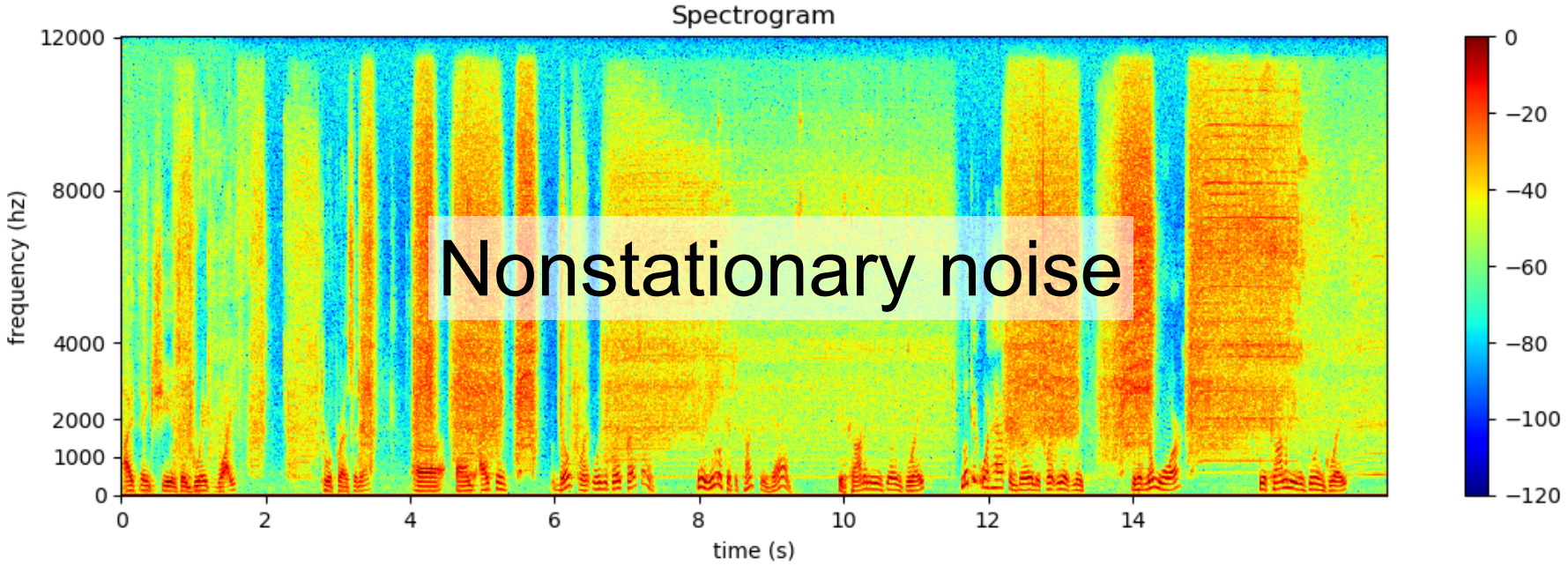


Noisy Input

DNN (proposed)

Minimum Tracking

Ideal Gains

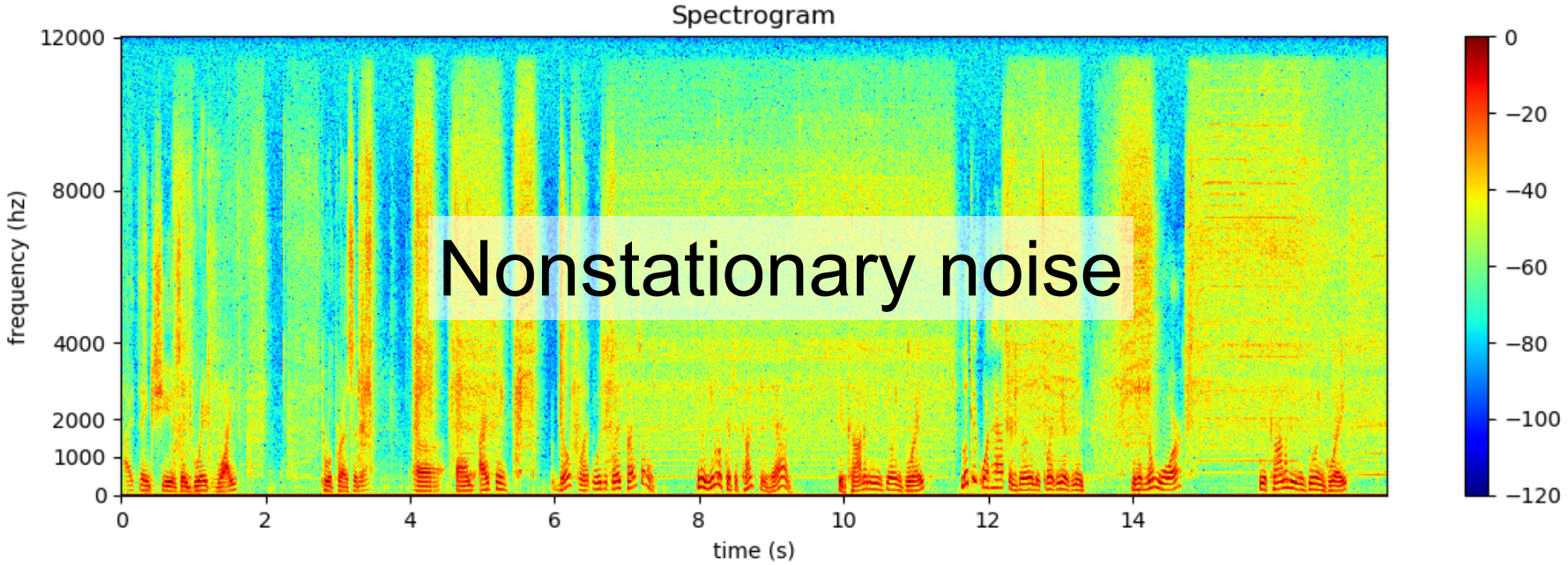


Noisy Input

DNN (proposed)

Minimum Tracking

Ideal Gains

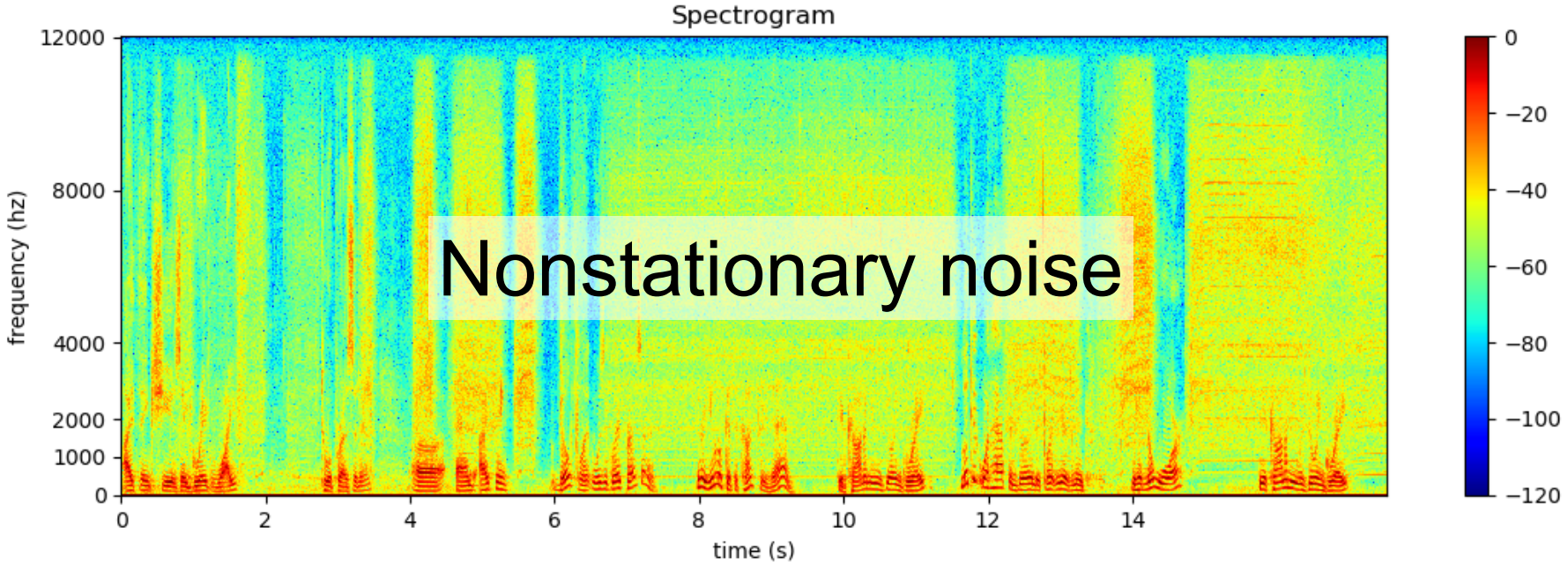


Noisy Input

DNN (proposed)

Minimum Tracking

Ideal Gains



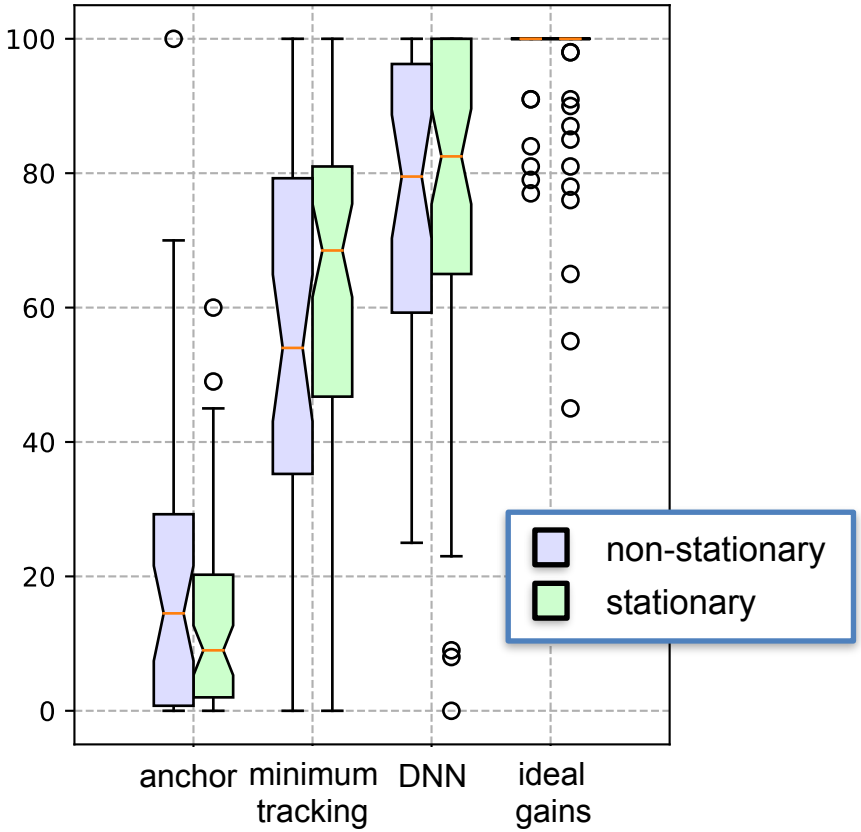
Noisy Input | DNN (proposed)

Minimum Tracking | Ideal Gains

Subjective Results

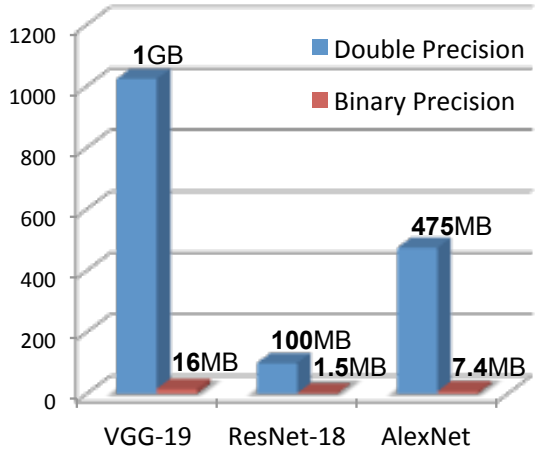
(MUSHRA, N=20)

- DNN as superior to recursive minimum tracking baseline
- Benefits especially for non-stationary signals.
- Still imperfect quality for low SNRs (phase distortion)



Deep Learning is too complex for hearing aids?

- DNNs are not dependent on floating point units.
- Even binary operations might be sufficient. (XNOR-Net [RO16])
- Potential for revival of analog computing [LY15]



[RO16]

[RO16] M. Rastegari et al., 2016, arxiv: 1603.05279
[LY15] Y. Lu et al, 2015, IEEE J. Solid State Circuits 50(1)

Summary: Hearing Aid Side-Conditions

- Limited processing power
  - ⇨ Likely not unsolvable in the future.
- Robustness required in any environment
  - ⇨ Promising, needs more evaluation.
- **Group delay > 10 ms objectionable to hearing aid wearers** [AT00]
  - ⇨ Total group delay = 8 ms.

[AT00]: J. Agnew and J. M. Thornton, JAAA (2000) 11:330-360

Thank you.

For more details:
IWAENC 2018 paper
https://arxiv.org/abs/1805.01198

Pattern Recognition Lab

sivantos the hearing company

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF ENGINEERING

Deep Neural Networks for Noise Reduction under Hearing Aid Side Conditions

M. Aubreville[1,2], K. Ehrensperger[1], T. Rosenkranz[2], B. Graf[2], H. Puder[2], A. Maier[1]

[1] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

[2] Sivantos GmbH, Erlangen, Germany

IHCON 2018, Lake Tahoe, CA, United States

Pattern Recognition Lab

sivantos the hearing company

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF ENGINEERING

Appendix

Normalization

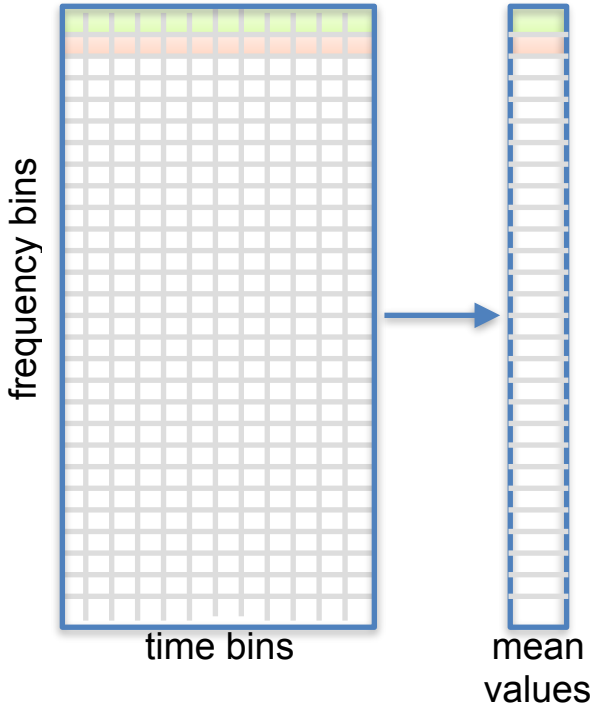
- Normalization across frequency:

$$X_{\text{norm}}(k,f) = X(k,f) - \frac{1}{\tau_1 + 1 + \tau_2} \sum_{k=-\tau_1}^{\tau_2} X(k,f)$$

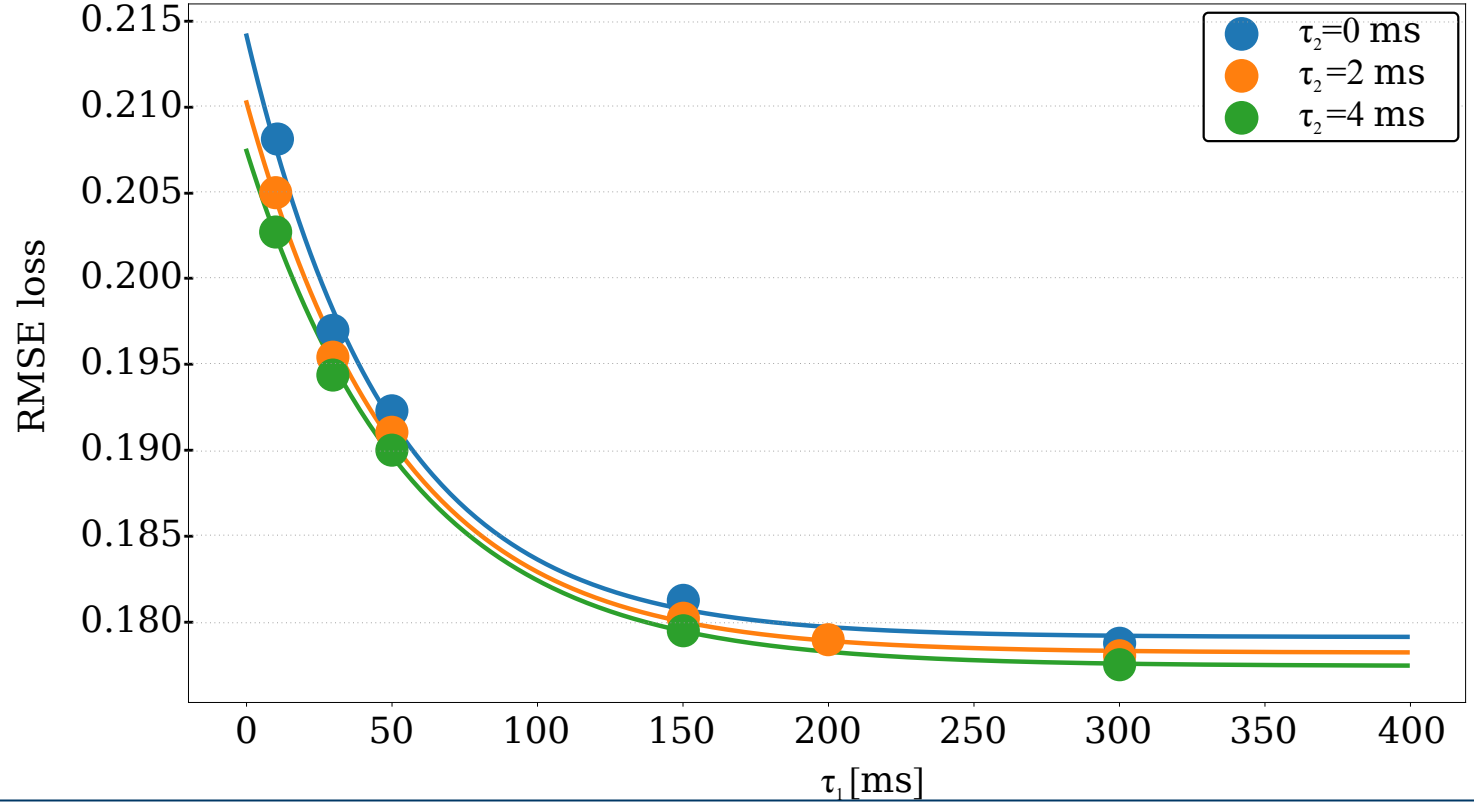
$X(k,f)$: Filterbank levels in log-scale

- Completely level independent

- Level information is fed as side-information to DNN



How much context is needed?



Context Size Influence on Gain Prediction