HörTech
Kompetenzzentrum für
Hörgeräte-Systemtechnik

Hearing Aid Research Data Set for Acoustic Environment Recognition (HEAR-DS)

ICASSP 2020

Andreas Hüwel *Kamil Adiloğlu* *Jörg-Hendrik Bach*

Hörtech gGmbH, Center of Competence for Hearing Systems
Marie-Curie-Str. 2, 26129, Oldenburg, Germany
email:{k.adiloglu, a.huewel, j.bach}@hoertech.de

This work was supported by the German Ministry of Education and Science (BMBF), FZK 02K16C202 “Audio-PSS”.

Hearing Aid Research Data Set for Acoustic Environment Recognition (HEAR-DS)

We propose a novel binaural data set

- Acoustic environment recognition
- Suitable for the needs of *hearing aids*
- Experimental validation by a group of baseline deep neural networks

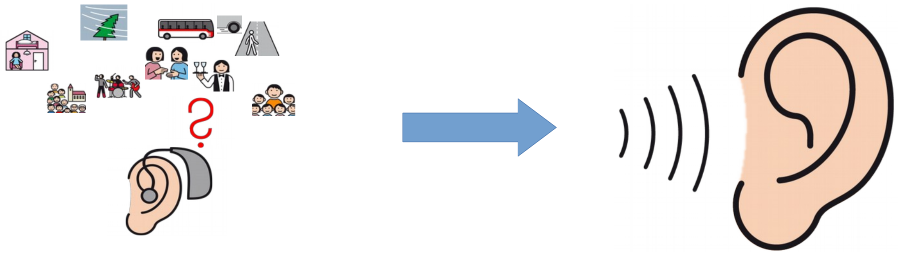
Current Situation

- Hearing aids provide several programs for different acoustic environments for enhancing the *quality and intelligibility of speech*.
- Reliable real-time recognition of current acoustic environment is essential.
- Limited computational resources:
  - Only simple, low-level features
  - compared with pre-defined threshold
  - to decide about the acoustic environment
- Even state-of-the-art hearing aids are limited in recognizing acoustic environments.

People can’t follow conversations in difficult environments

Machine Learning towards Hearing Aids

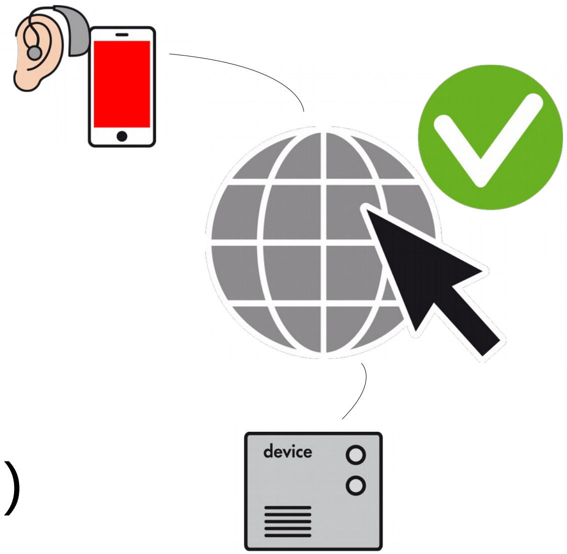
With machine learning, different noisy acoustic environments can be recognized



and then optimally suppressed, which in return yields a better intelligibility and quality of speech.

Internet of Things (IoT) approach

- Connect many wearers with each other
- Computational burden,
  e.g. training of neural networks,
  is delegated to a *cloud computing system*.
- Hearing Aid
  - performing only the recognition (not the training)
  - using the trained model only in forward mode
  - feasible challenge even inside computational limits of a hearing aid.

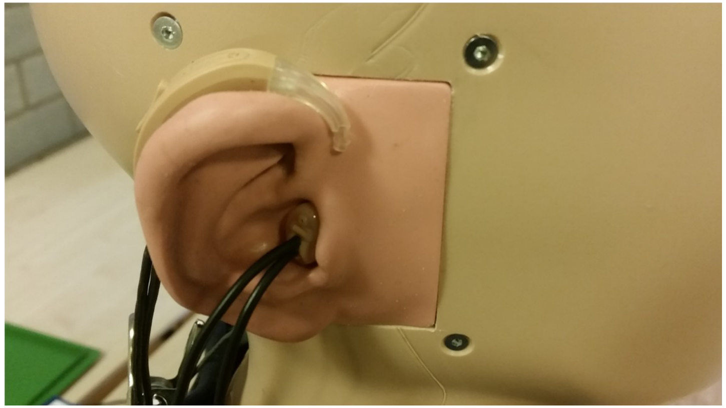


Training Data

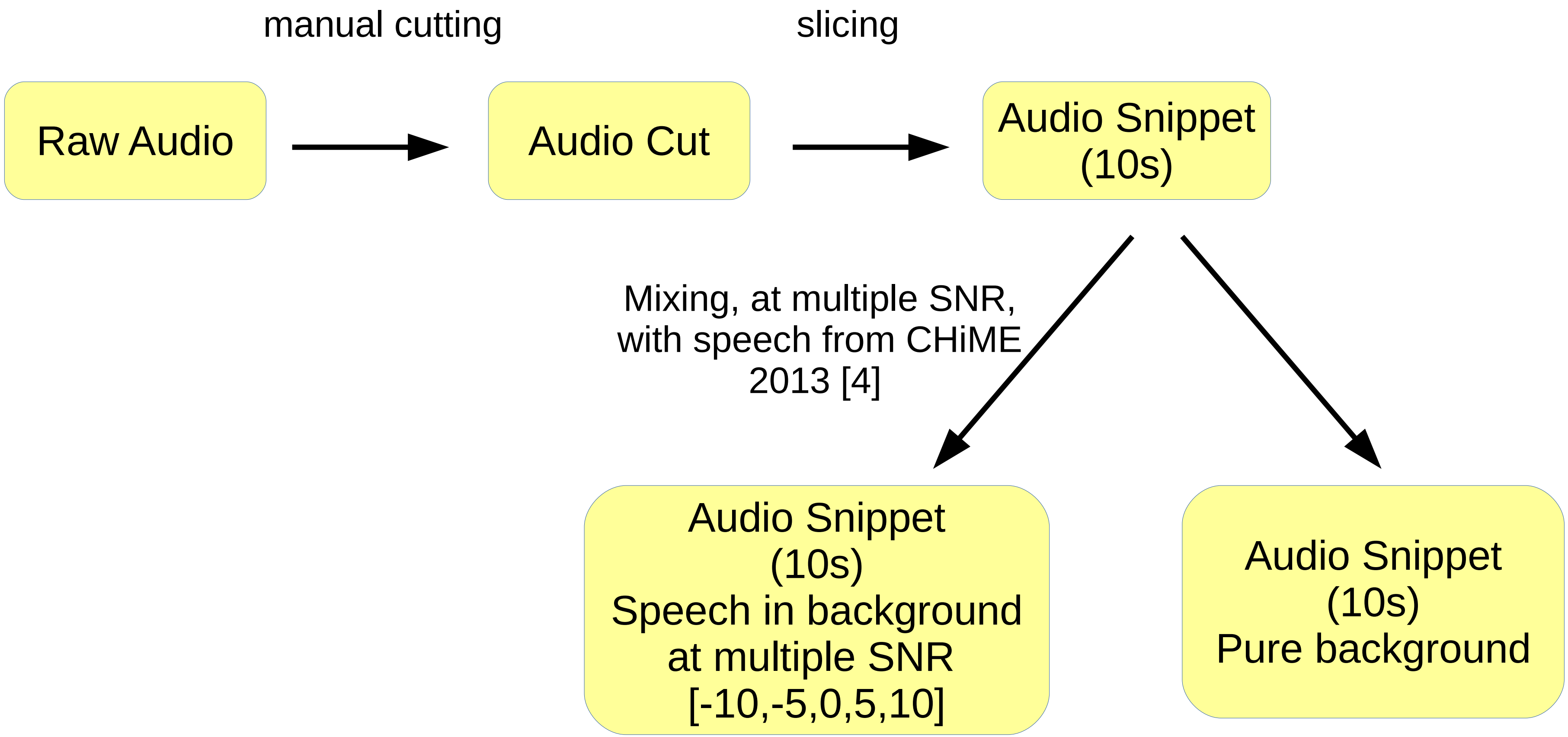
- To train such models, a large training data set is required.
- Existing data sets
  - DCASE [1], MsoS [2], LITIS [3], ChiME [4,5], MIREX [6], Freesound [8] etc.
  - define label scenes according to the ***location***
- Hearing aids need to group similar acoustic features together as ***acoustic environments***
- *HEAR-DS:* suitable for the needs of hearing aids

Audio Recording

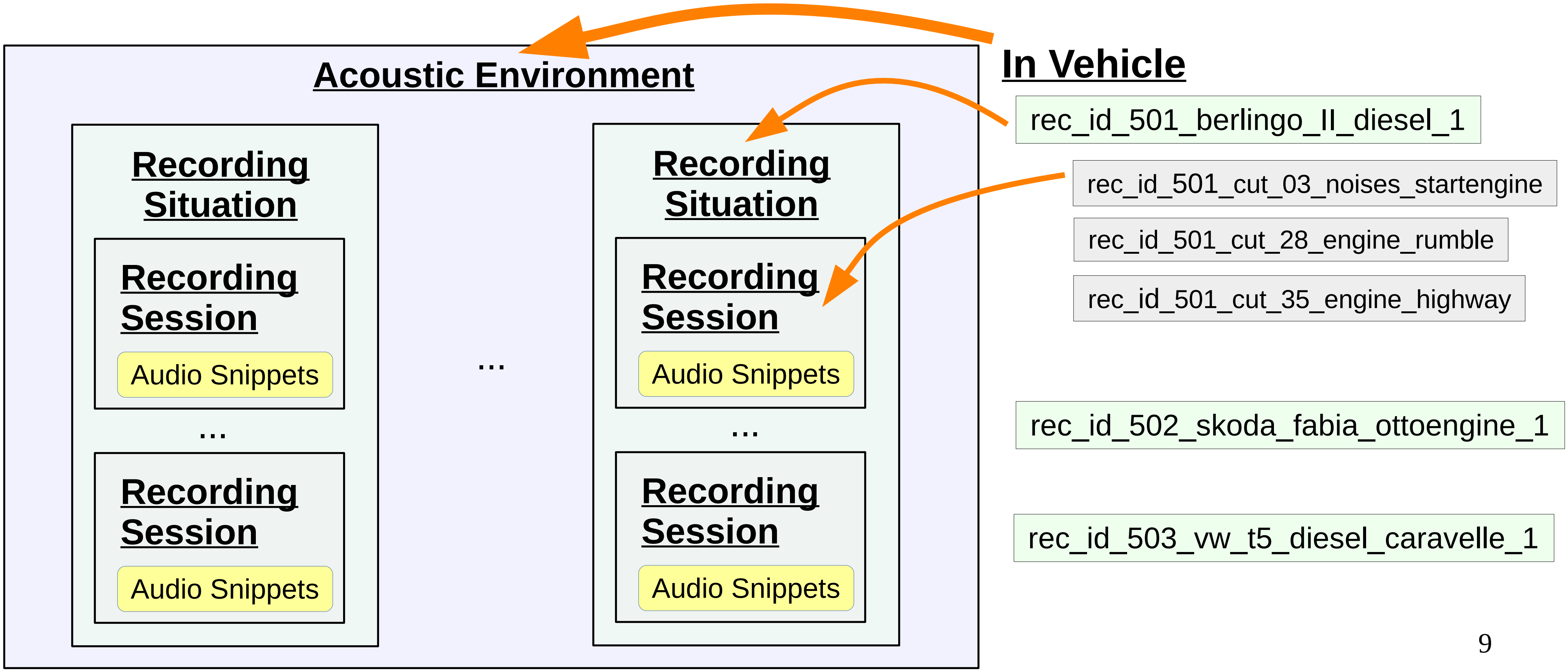
- binaural recordings with hearing aid on Artificial Head
  - with adjustable ear canals (DADEC [9]) equipped with G.R.A.S. KB 1065/1066 Pinnae
  - ITC 2 mics (L/R)
  - BTE 4 mics (L/R, each front/rear)
- Pre-Amp (for each mic)
  - with fixed amplification factor 100
- Focusrite Scarlet 18i6 soundcard
  - at 48 kHz in 32-bit PCM



Audio Material

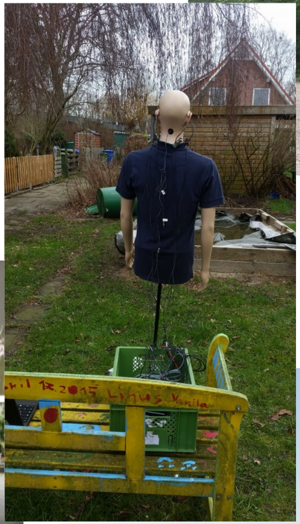


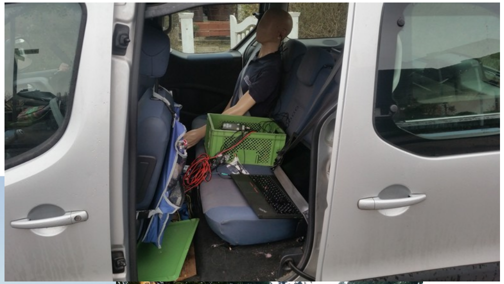
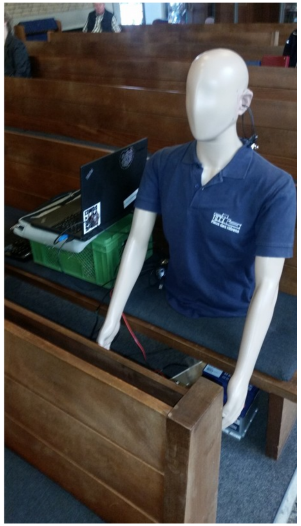
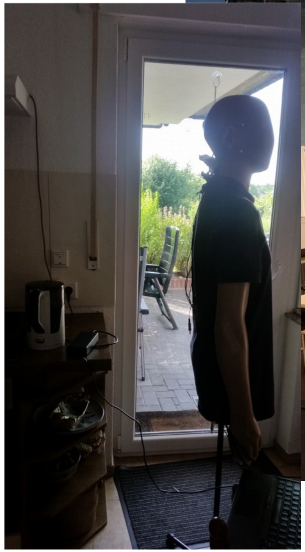
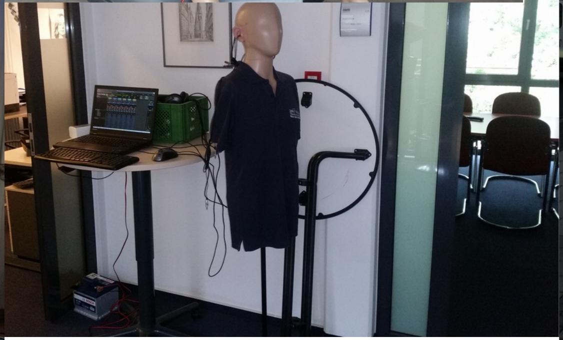
Structuring for Machine Learning









HörTech
Kompetenzzentrum für Hörgeräte-Systemtechnik

HEAR-DS Environments

Speech			
Cocktail party	667		
Interfering speakers	1481		
Background		**Speech in background**	
In traffic	530	*Speech in traffic*	470
In vehicle	584	*Speech in vehicle*	511
Music	1496	*Speech in music*	1495
Quiet indoors	525	*Speech in qu. indoors*	426
Reverberant env.	315	*Speech in reverb. env.*	692
Wind turbulence	595	*Speech in wind turb.*	439

HörTech
Kompetenzzentrum für
Hörgeräte-Systemtechnik

HEAR-DS Environments

Interfering speakers: CHiME 2018 [5]

Speech for mixing: CHiME 2013 [4]

Music: GTZAN [7], resampled to 48kHz convolved with binaural head-related transfer function (Kayser [10])

Speech			
…ail party	667		
Interfering speakers	1481		
Background		**Speech in background**	
In traffic	530	*Speech in traffic*	470
In vehicle	584	*Speech in vehicle*	511
Music	1496	*Speech in music*	1495
…ors	525	*Speech in qu. indoors*	426
…nv.	315	*Speech in reverb. env.*	692
…ce	595	*Speech in wind turb.*	439

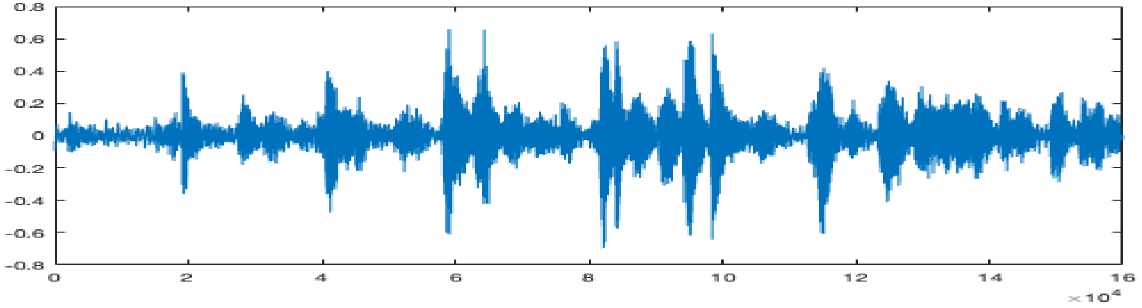
Validation Experiment

- Goal: Show separability of acoustic environments by deep neural networks
- Challenge:
  - lightweight networks
  - still reach good recognition rates
- series of classification experiments with decreasing complex deep neural networks

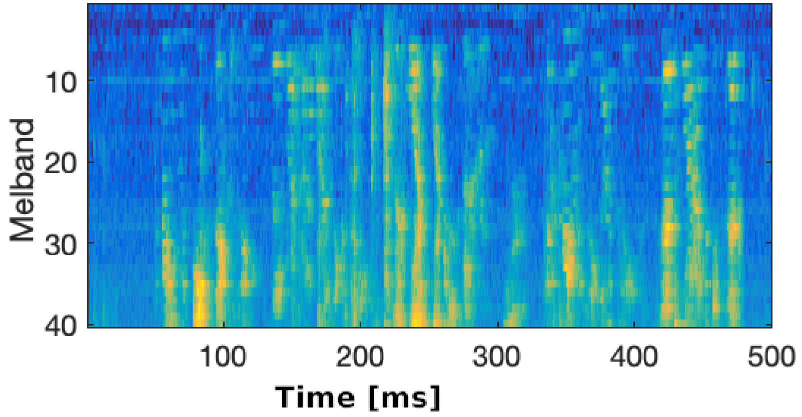
Challenge

- Streamlined small
  but still accurate DNNs
- optimized for low computational resources
- for real-time capable applications
- toward hearing aids

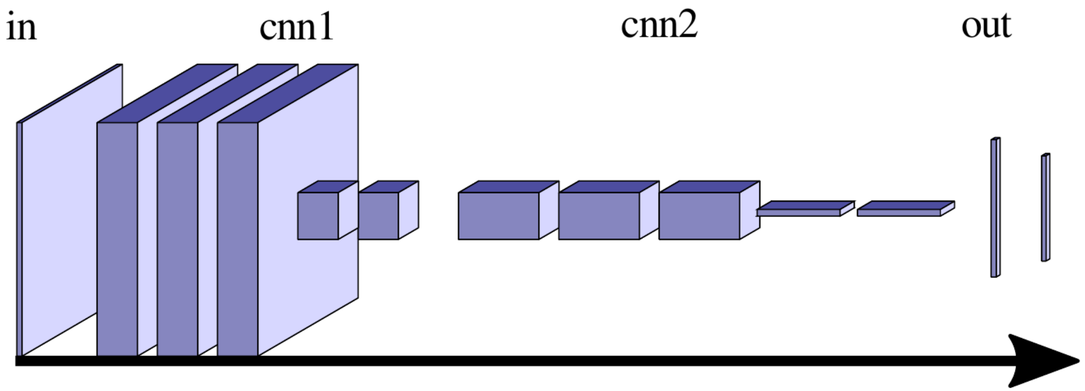
Feature Extraction



$$\mathrm{logmel}(\mathbf{x}_{j,fn}) = 20\log_{10}(\mathbf{M}|\mathbf{x}_{j,fn}|)$$



Network Architecture: Topology



Optimizer:
torch.optim.SGD
+
nn.CrossEntropyLoss

cnn 1|2
1) nn.Conv2d()
2) nn.BatchNorm2d()
3) nn.ReLU()
4) nn.MaxPool2d()
5) nn.Dropout()

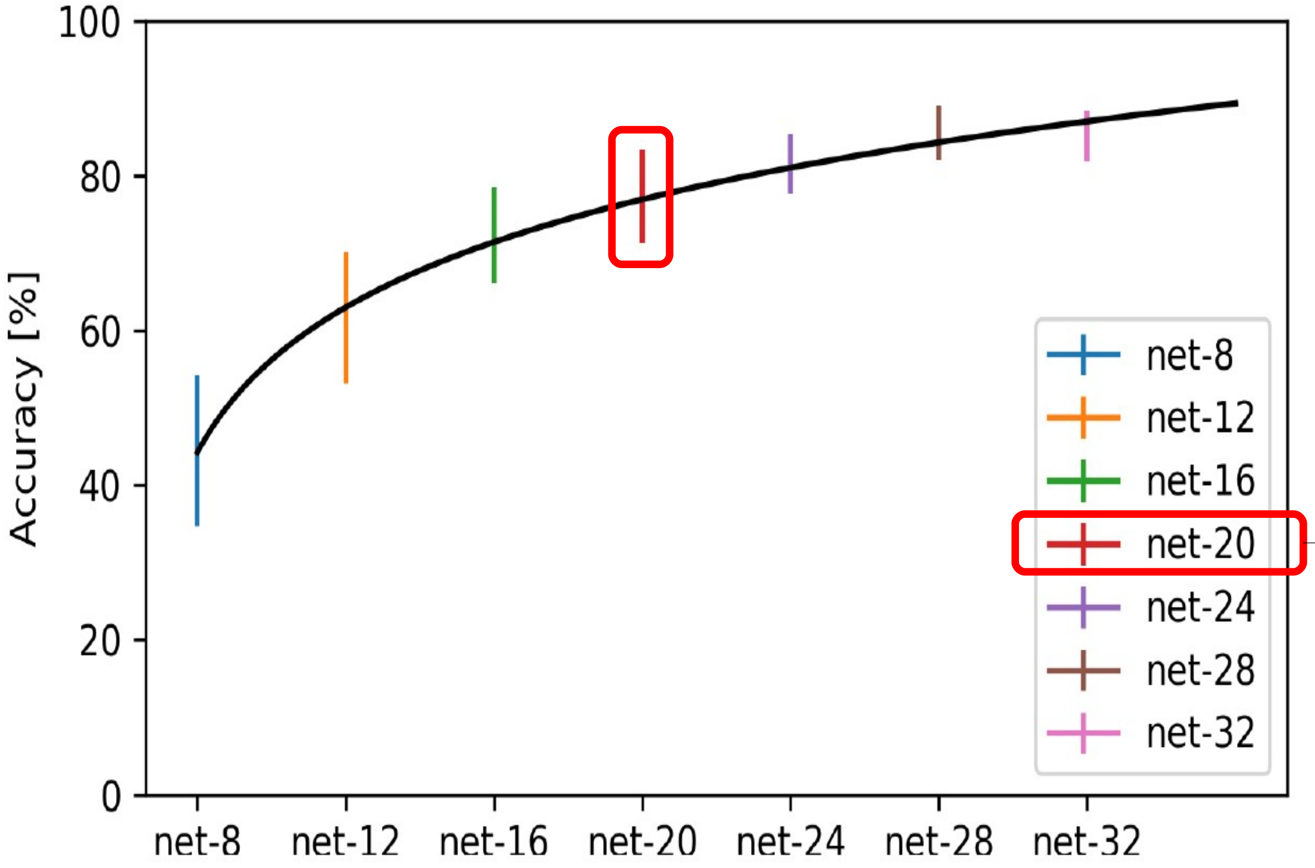
out
1) nn.Linear()
2) nn.Dropout()
3) nn.Linear()

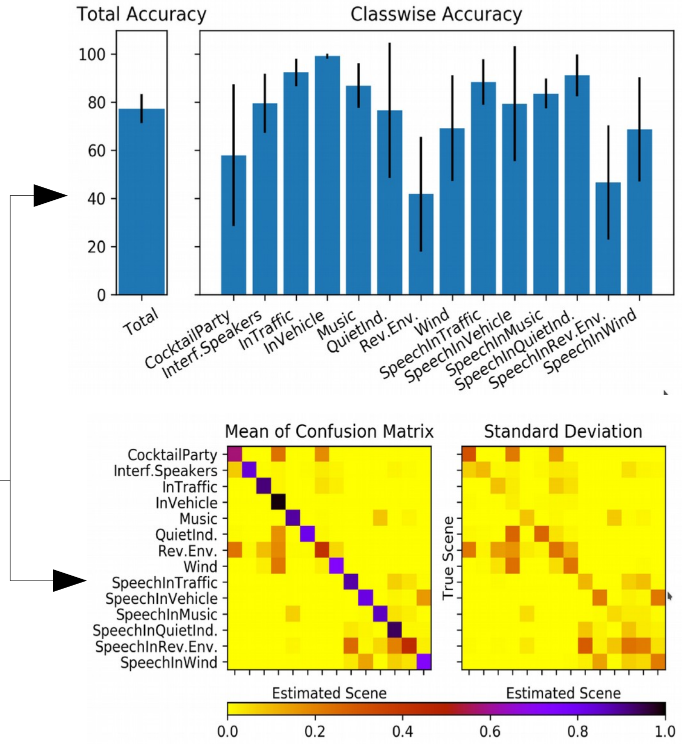
HörTech
Kompetenzzentrum für
Hörgeräte-Systemtechnik

Decreasing Complexity of Network Architectures

Network	$\mathbf{CNN_1}$	$\mathbf{CNN_2}$	**FC**
net-32	32	64	100
net-28	28	56	87
net-24	24	48	75
net-20	20	40	63
net-16	16	32	50
net-12	12	24	37
net-8	8	16	25

Experiment Results





Live Evaluation System

- NUK mini PC
- C++
- importing pre-trained PyTorch-model
- audio induced via loudspeaker over hearing aid
- *Net-32* takes < 0.4s to recognize 10s audio

More in the show and tell session ICASSP 2020, Thu 7. May 11:30

Conclusions

- Provided results show
  - validity of the data set
  - the data set can be classified
  - live audio recognition on a mini PC
- Further research needed
  - HEAR-DS enables researchers to test algorithms on different acoustic environments
  - optimize DNNs for hearing aids
    - Robustness
    - Real-time
    - limited computational capability
- **Make use of HEAR-DS [11]**
  - We provide the Data
    anything that can be made free is made free

https://www.hoertech.de/en/research/open-tools-for-science.html

References

- DCase [1] Annamaria Mesaros et al., “A multi-device dataset for urban acoustic scene classification,” in Detection and Classification of Acoustic Scenes and Events 2018 Workshop (DCASE2018), 2018.
- MsoS [2] Making Sense of Sounds Project, “Making sense of sounds challenge,” http://cvssp.org/projects/making_ sense_of_sounds/site/challenge/, 2018, accessed 2019-10-21.
- LITIS [3] Alain Rakotomamonjy and Gilles Gasso, “Histogram of gradients of time-frequency representations for audio scene detection,” IEEE/ACM Trans. Audio, Speech, and Language Processing, vol. 23-1, pp. 142–153, 2015.
- CHime 2013 [4] Emmanuel Vincent et al., “The second CHiME speech separation and recognition challenge: Datasets, tasks and baselines,” in IEEE International Conference on Acoustics, Speech, and Signal Processing, Vancouver, Canada, 2013.
- Chime 2018 [5] Jon Barker et al., “The fifth CHiME speech separation and recognition challenge: Dataset, task and baselines,” in Interspeech, Hayderabad, India, 2018.
- MIREX challenge [6] J. Stephen Downie et al., “The music information retrieval evaluation exchange: Some observations and insights,” in Advances in Music Information Retrieval, vol. 274, pp. 93–115. Springer, 2010.
- GTZAN [7] Giannis Tzanetakis and Perry Cook, “Musical genre classification of audio signals,” IEEE Transactions on Speech and Audio Processing, vol. 10, no. 5, pp. 293–302, July 2002.
- Freesound [8] Frederic Font et al., “Freesound technical demo,” in ACM International Conference on Multimedia (MM’13), Barcelona, Spain, 21/10/2013 2013, ACM, pp. 411–412, ACM.
- DADEC [9] Marko Hiipakka et al., “Modeling the external ear acoustics for insert headphone usage,” Journal of the Audio Engineering Society, vol. 58-4, pp. 269–281, 2010.
- Kayser [10] Hendrik Kayser et al., “Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses,” EURASIP Journal on Advances in Signal Processing, vol. 2009, no. 1, pp. 298605, 2009.
- HEAR-DS [11] HörTech gGmbH, “Open Research Tools For Science, HEAR- Data Set,” https://www.hoertech.de/en/research/open-tools-for-science.html, 2019, accessed 2019-10-22.