Deep Neural Networks for speech enhancement in noise

AUTHORS: Lars Bramsløw[1], Gaurav Naithani[2], Tuomas Virtanen[2]

[1]Eriksholm Research Centre, [2]Tampere University

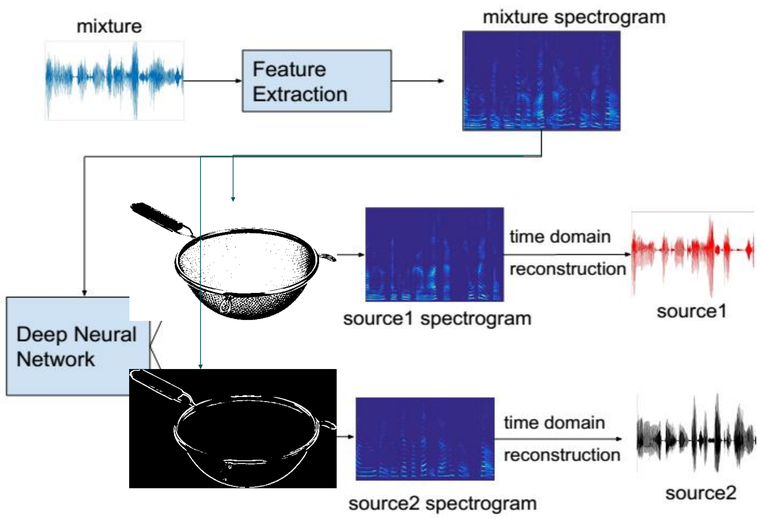
Deep neural networks (DNN) have demonstrated substantial user benefits for voice-on-voice enhancement (Healy et al 2017, Bramsløw et al, 2018). In the latter paper, two known talkers were separated, using different types of low-latency DNN algorithms.

The present work uses similar DNN architectures for enhancing speech from more common noise types: a party noise and a shopping centre ambient noise.

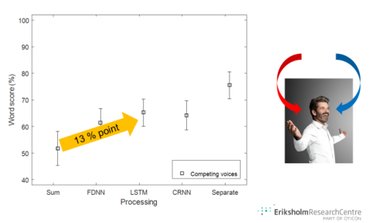
Background

Good results for two talkers (Bramsløw et al, 2018; SPIN 2018)

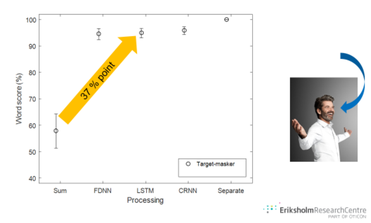
DNN provided 13%-point benefit in competing known voices with dual targets presented binaurally, and 37%-point benefit if chosen target presented to both ears (Bramsløw et al, 2018).



Speech segregation results



Speech separation results



Test design

Processing:
1. Sum (= input)
2. FDNN known voice
3. LSTM known voice
4. LSTM unknown voice
5. LSTM unknown voice + multi resolution mask
6. LSTM unknown voice + phase sensitive mask
7. Ideal ratio mask

Max. 20 dB attenuation (except 7.)
Roughly 3.5 mio weights
5 ms latency (Naithani et al, 2017)

Noise:
Party (P1), shopping (S1).

Listeners:
21 moderate, sloping losses, CAMEq

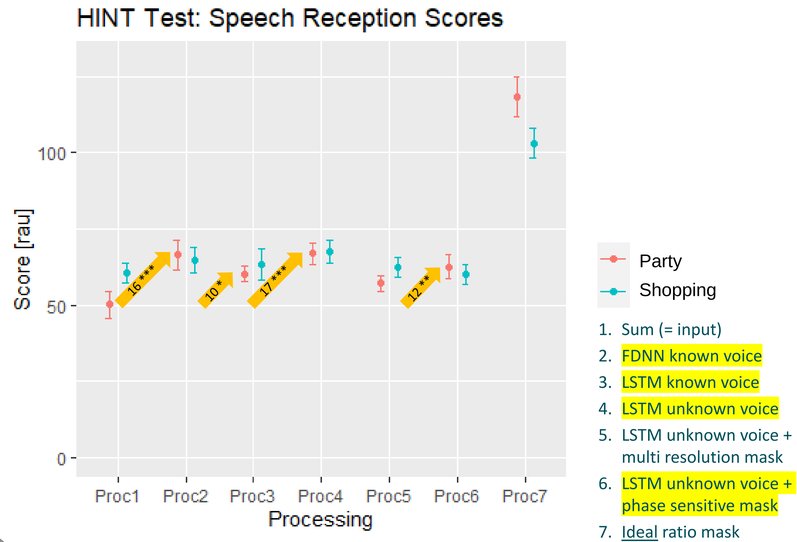
Speech for training and test

Danish HINT sentences
- M1-M6, F1-F6
- 200 – 260 sentences ~ 6 min

Target talkers:
- M1, M2, F1, F3
- Known voice: train on these (test other sentences)
- Unknown voice: do not train on these (test all sentences)

Noise from the 'ICRA natural sound library'
- P1: Party noise
  - train at -3..+3 dB SNR, test at +0 dB
- S1: Shopping center noise
  - train at -3..+3 dB, test at +0 dB.

Test items ≠Training items



Conclusions

DNN benefit in party noise
- ~17 %-point (~1.7 dB)
- known voice FDNN
- unknown voice LSTM

No benefit in shopping centre
- Less modulated = less glimpses

Unknown noise remains unsolved.

Information

Lars Bramsløw
Eriksholm Research Centre
labw@eriksholm.com

Eriksholm Research Centre
Rørtangvej 20
DK - 3070 Snekkersten
Phone +45 4829 8900

Read more at:
https://www.eriksholm.com/publication/

In collaboration with:
Computing Sciences,
Tampere University, Finland

References: Bramsløw, L., Naithani, G., Hafez, A., Barker, T., Pontoppidan, N. H., and Virtanen, T. (2018). "Improving competing voices segregation for hearing impaired listeners using a low-latency deep neural network algorithm," J. Acoust. Soc. Am., 144, 172–185. ; Naithani, G., Barker, T., Parascandolo, G., Bramslow, L., Pontoppidan, N. H., and Virtanen, T. (2017). "Low latency sound source separation using convolutional recurrent neural networks," 2017 IEEE Work. Appl. Signal Process. to Audio Acoust., IEEE, New Paltz, NY, 71–75. ; Healy, E. W., Delfarah, M., Vasko, J. L., Carter, B. L., and Wang, D. (2017). "An algorithm to increase intelligibility for hearing-impaired listeners in the presence of a competing talker," J. Acoust. Soc. Am., 141, 4230–4239.

TAMPERE UNIVERSITY OF TECHNOLOGY

EriksholmResearchCentre
PART OF OTICON