# The Effect of DNN-based Voice Segregation on the Selective Attention

AUTHORS: Emina Alickovic[1,2], Lars Bramsløw[1], Umaer Hanif[1], Rikke Rossing[1], Gaurav Naithani[3], Tuomas Virtanen[3], Carina Graversen[1], Dorothea Wendt[1,4], Thomas Lunner[1,2,4,5], Niels Henrik Pontoppidan[1]

[1]Eriksholm Research Centre, Denmark, [2]Department of Electrical Engineering, Linkoping University, Sweden, [3]Laboratory of Signal Processing, Tampere University of Technology, Tampere, Finland, [4]Department of Health Technology, Technical University of Denmark, Denmark [5]Department of Behavioral Sciences and Learning, Linkoping University, Sweden

## Introduction

### Background

- Deep neural networks (DNNs) have a high potential for improving voice segregation and speech intelligibility in multiple-voice environments [1].
- Neural responses can be decoded to segregate the voice of a listener's interest by providing separated access to each speaker via separate presentation [2].

### Motivation

This study investigates the effect of **a low-latency DNN** algorithm for separating two competing voices on selective attention for HA users.

### Research question

*Can **auditory attention decoding (AAD)** methods* [3] *be used to evaluate the effect of* **a low-latency DNN speech separation** algorithm in hearing-aid *users?*

## Experiment

### Participants

**15 HI** subjects with an average age of 70 ± 12.

### EEG data Acquisition

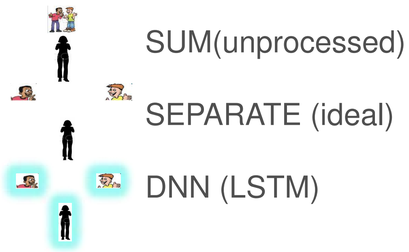
64 channels of scalp EEG data (10/20 system) were recorded using the Biosemi ActiveTwo system.

### Stimuli

- **Danish** continuous speech material (100 minutes).
- 1 male + 1 female.
- Scaled to same RMS.
- **DNN** procedure:
  - DNN Training: 20 min
  - DNN Validation: 10 min
  - Available for EEG test: 70 min

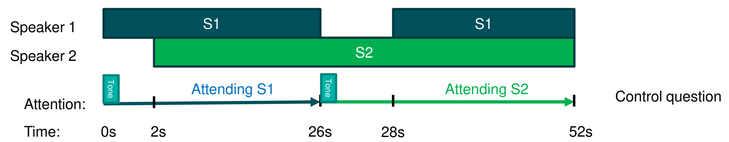
## Experiment design

### Test design:

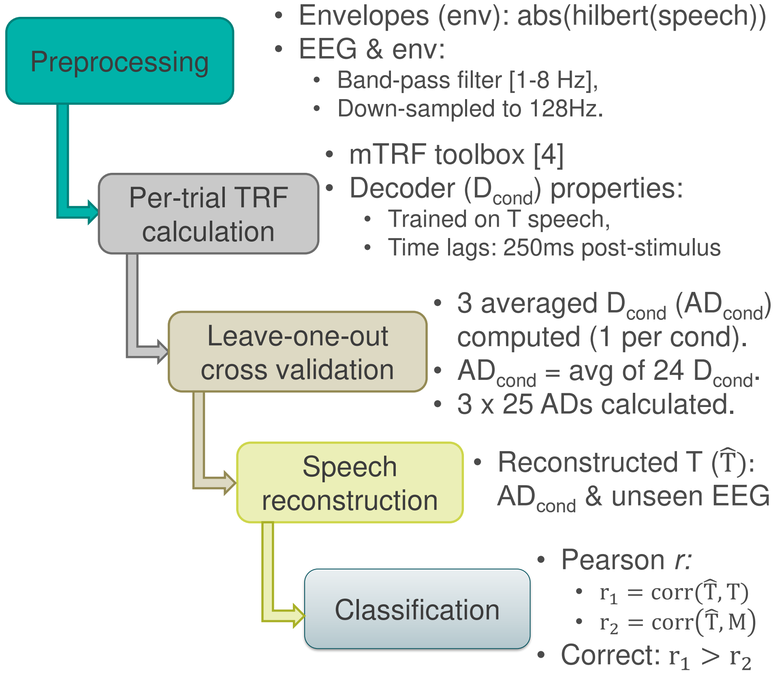



- Total = 3 conditions
- 25 replications = 75 recordings per subject

### Task design:



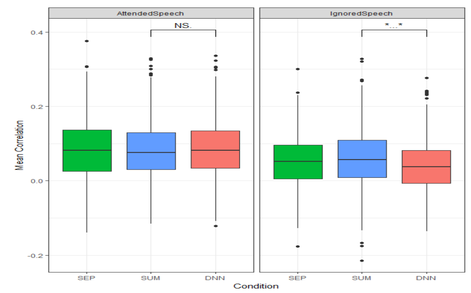

## Data analysis method




## Results

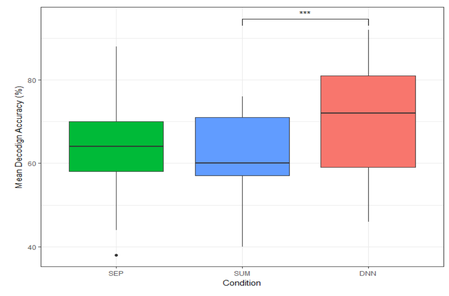
### Attention decoding

#### Mean correlations




DNN condition had:
- Sig lower masker correlations when compared to the SUM condition (p <0.0005).
- Non-sig lower masker correlation when compared to the SEPARATE condition.

#### Attention decoding accuracy




- DNN better decoded than SUM (p < 0.005)
- But not better than SEPARATE
- The mask may be helpful

## Conclusion

Data analysis showed that the DNN-based voice segregation had a significant effect on selective attention, demonstrating the potential of the low-latency DNN algorithms.

**Information**
Emina Alickovic
Eriksholm Research Centre
eali@eriksholm.com

Read more at:
www.eriksholm.com:

**References** [1] Bramsløw, Lars, et al. "Improving competing voices segregation for hearing impaired listeners using a low-latency deep neural network algorithm." The Journal of the Acoustical Society of America 144.1 (2018): 172-185. [2] O'sullivan, James A., et al. "Attentional selection in a cocktail party environment can be decoded from single-trial EEG." Cerebral Cortex 25.7 (2014): 1697-1706. [3] Alickovic, Emina, et al. "A Tutorial on Auditory Attention Identification Methods." Frontiers in Neuroscience 13 (2019): 153. [4] Crosse, Michael J., et al. "The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli." Frontiers in human neuroscience 10 (2016): 604.

EriksholmResearchCentre
PART OF OTICON