Towards auditory attention decoding with noise-tagging: A pilot study

Hanneke Scheppink, Sara Ahmadi, Peter Desain, Michael Tangermann & Jordy Thielen

Radboud University, Donders Institute for Brain, Cognition and Behaviour, Machine Learning and Neural Computation, Data-Driven Neurotechnology Lab, Nijmegen, the Netherlands

INTRODUCTION

Background

Auditory attention decoding (AAD) aims to **extract from brain activity the attended speaker** amidst candidate speakers, offering promising applications for neuro-steered hearing devices and brain-computer interfacing[1]. Current state-of-the-art algorithms for AAD can do so with a mean accuracy of about 85 % using 30 s decision windows.

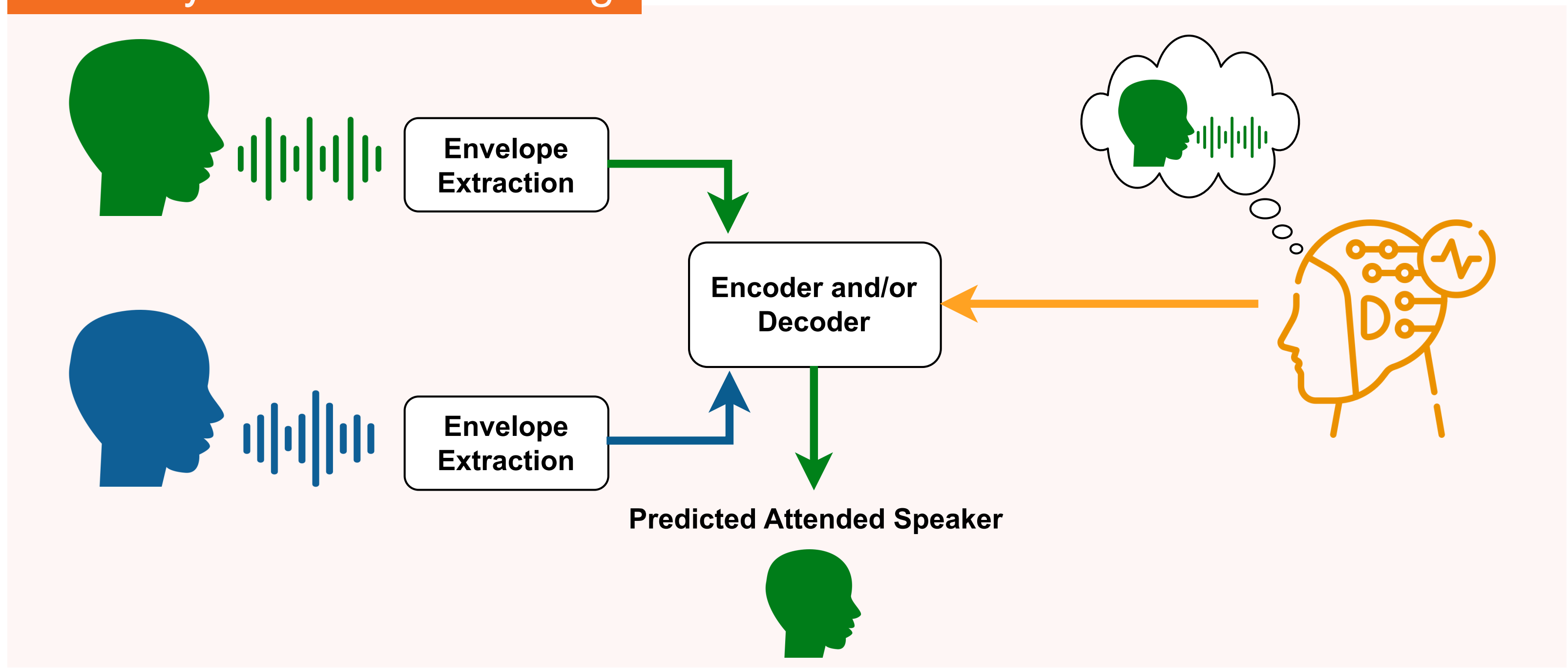
Challenge

For real-world scenarios, a fast and accurate speaker detection is crucial to **keep up with the quick dynamics of natural language and turn-taking**. Unfortunately, when decreasing the decision window length to about 10 s, the accuracy quickly drops to below 80 %. This poses a significant limitation for AAD.

Approach

BCIs based on the code-modulated visual evoked potential (c-VEP), the response to rapid pseudo-random visual stimulation, have recently shown incredible performance surpassing previous records[2]. In this study, we aim to **leverage the code-modulated stimulus protocol to improve speaker identification in AAD**.

Auditory Attention Decoding



EXPERIMENT

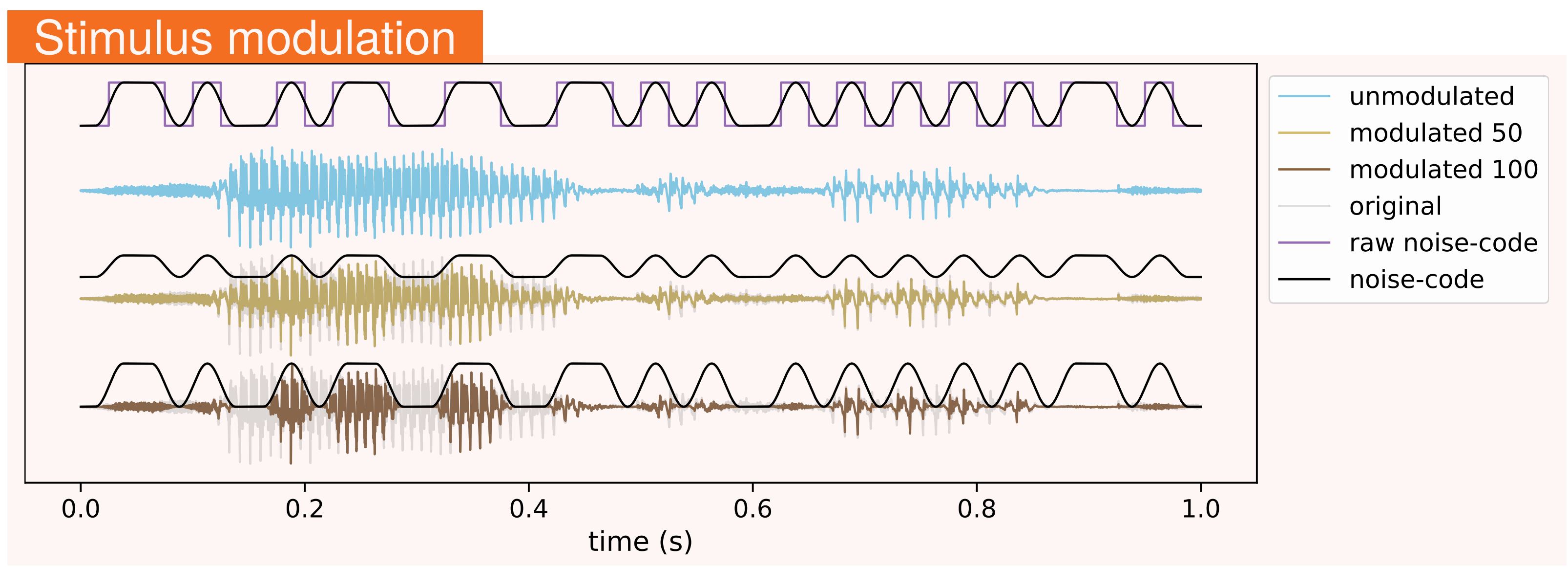
Recording

- 64-channel EEG
- BrainProducts BrainAmp
- 500 Hz sampling rate
- 5 participants
- 50 Hz notch filter
- 1-20 Hz bandpass

Stimulus modulation and presentation

- Two Dutch stories (6.5 min)[3]
- Five modulation depths (md)
- 0 (unmodulated), 50, 70, 90, 100 md
- Temporal muting of the audio signal
- Noise-codes: two modulated Gold codes[4]
- Smoothen sharp edges binary code
- Multiply smooth codes with audio signal
- Presented sequentially

Stimulus modulation



Procedure

- Five runs; one for each condition
- One run; two trials (6.5 min); one per story
- First story to left ear, silence on right, followed by second story to right with silence on left

ANALYSIS

Canonical correlation analysis (CCA)

CCA is a state-of-the-art model in both AAD[1] and BCI[2]. Commonly, it learns a spatial filter $\mathbf{w} \in \mathbb{R}^C$ for C channels and a temporal filter $\mathbf{r} \in \mathbb{R}^L$ of L samples, using training data $\mathbf{X} \in \mathbb{R}^{C \times T}$ of T samples and stimulus information $\mathbf{Z} \in \mathbb{R}^{L \times T}$, optimizing the correlation ρ:

$$\underset{\mathbf{w},\mathbf{r}}{\arg\max}\, \rho(\mathbf{w}^\top \mathbf{X}, \mathbf{r}^\top \mathbf{Z}_i) \tag{1}$$

Determining the attended target symbol $\hat{y}$ is performed by maximizing the correlation ρ:

$$\hat{y} = \underset{i}{\arg\max}\, \rho(\mathbf{w}^\top \mathbf{X}, \mathbf{r}^\top \mathbf{Z}_i) \tag{2}$$

Envelope CCA (eCCA)

The synchronization between the listener's brain signals and the attended speech envelope is stronger than with the ignored speech envelope[1]. Therefore, the AAD field commonly uses the envelope, for instance from a Gammatone filter[5]. Hence, for eCCA, $\mathbf{Z}_i = \mathbf{E}_i$ where $\mathbf{E}_i$ is the envelope matrix of the i-th speaker, with L delayed versions.

Reconvolution CCA (rCCA)

The rCCA method originates from the c-VEP field and models the response to a sequence of events as the linear summation of the responses to the individual events[4]. It denotes the event onsets, duration, and overlap in a Toeplitz matrix $\mathbf{M}_i$. This matrix is multiplied with the envelope to incorporate the intensity of the modulated audio played, hence, $\mathbf{Z}_i = \mathbf{M}_i$.

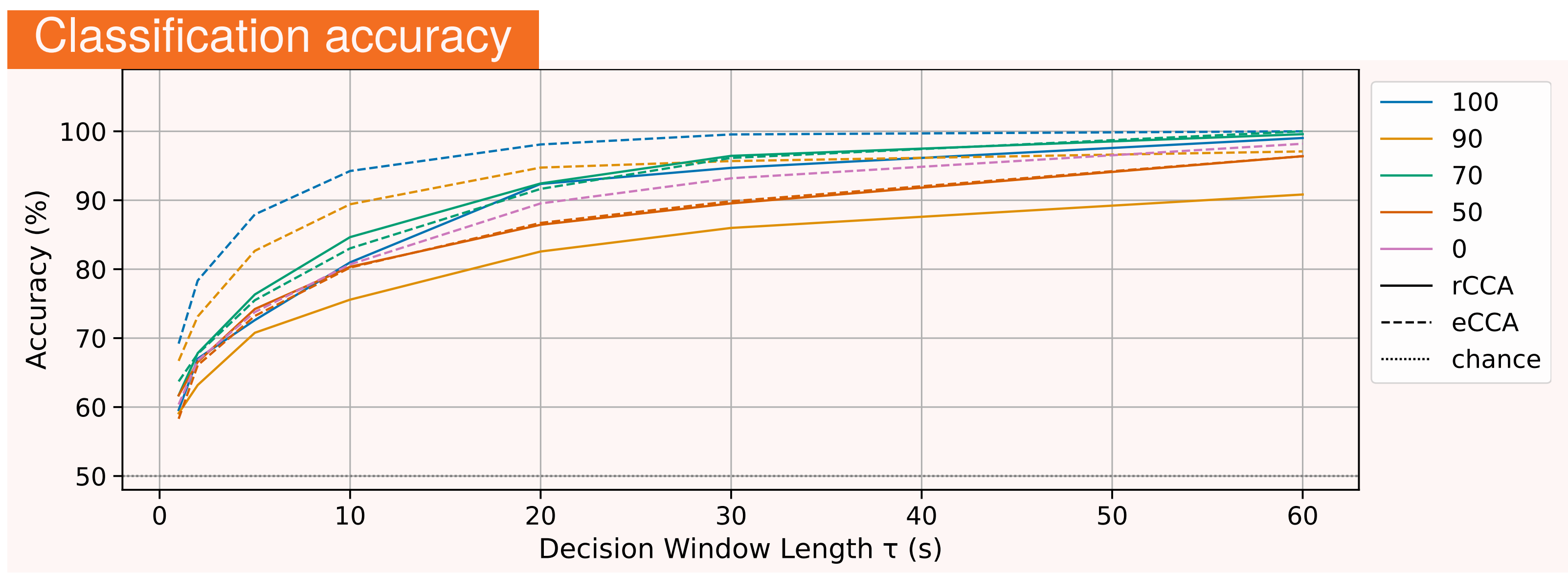
RESULTS

Classification accuracy

A number of observations were made:

- Overall, eCCA reaches higher performances than rCCA.
- 0 md (full intensity) performs lower than 70 md, 90 md, and 100 md, but not lower than 50 md.
- With 70 md, already an accuracy of 85 % is obtained after 10 s, which increases to a perfect decoding at 60 s.

Classification accuracy



Classification accuracy

depth [md]	method	1 s	10 s	30 s	60 s
0	eCCA	60.4	80.7	93.2	98.2
50	eCCA	58.3	80.2	**89.9**	96.4
	rCCA	**61.7**	**80.4**	89.6	96.4
70	eCCA	**63.7**	83.0	96.1	**100.0**
	rCCA	61.7	**84.7**	**96.4**	99.6
90	eCCA	**66.7**	**89.4**	**95.7**	**97.1**
	rCCA	59.1	75.6	86.0	90.8
100	eCCA	**69.2**	**94.2**	**99.6**	**100.0**
	rCCA	59.6	81.0	94.7	99.0

DISCUSSION

Realizing auditory noise-tagging

Our pilot study demonstrated that **incorporating noise information into speech signals can enhance decoding performance**. This improvement may result from introducing additional decodable information and the decorrelation of potentially correlated audio sources. These findings represent an initial step toward applying noise tagging in the auditory modality and advancing auditory attention decoding.

OPEN QUESTIONS

Towards parallel stimulation

- Two key limitations:
  - Small sample size
  - Sequential stimulation
- Valuable insights into feasibility c-AEP
- First step of c-AEP for AAD
- However, parallel stimulation is needed for more practical AAD application

Optimizing CCA

- Performance rCCA on-par or slightly lower than eCCA
- eCCA emphasizes higher-order brain activity (speech envelopes)
- rCCA may be more attuned to early sensory responses (noise-codes)
- eCCA optimized for AAD domain, rCCA originally for c-VEP domain
- Future work: optimize rCCA for AAD domain

Effect of amplitude modulation on speech intelligibility

- Determining optimal modulation depth
- 100 md yielded highest performance
  - Audio more unintelligible
  - Uncomfortable to listen to
- 70 md outperformed no modulation (0 md)
  - Less audio distortion than 100 md
- Future work:
  - Behavioural study to explore threshold at which modulation becomes inaudible
  - Study decoding performance at threshold
  - Optimizing noise-codes to preserve speech characteristics

Generalization to other domains

- Noise-codes can be embedded in any audio signal (e.g. music)
- Could make signal more orthogonal, thereby enhancing decoding
- Noise-codes in tactile domain; new possibilities for a diverse range of users

REFERENCES

[1] Geirnaert et al. (2021) *IEEE Signal Proc Mag* doi:10.1109/MSP.2021.3075932

[2] Martinez-Cagigal et al. (2021) *J Neural Eng* doi:10.1088/1741-2552/ac38cf

[3] Das et al. (2016) *J Neural Eng* doi:10.1088/1741-2560/13/5/056014

[4] Thielen et al. (2015) *PLOS ONE* doi:10.1371/journal.pone.0133797

[5] Biesmans et al. (2017) *IEEE T Neur Sys Reh* doi:10.1109/TNSRE.2016.2571900

E-mail: jordy.thielen@donders.ru.nl
Twitter: https://twitter.com/ThielenJordy
Lab: https://neurotechlab.socsci.ru.nl/

Poster PDF:

Presented at the 9th Graz Brain-Computer Interface Conference 2024 in Graz, Austria on Thursday 12th September 2024