# On Least-squares-based Auditory Attention Decoding with Individual Neural Latency Compensation

Daniel Neudek[1], Anil Nagathil[1], Alexander Wißmann[1], Stephan Getzmann[2], Rainer Martin[1]

[1] *Ruhr-Universität Bochum, Institute of Communication Acoustics, Bochum, Email: firstname.lastname@rub.de*

[2] *Leibniz Research Centre for Working Environment and Human Factors - IfADo, Dortmund, Email: getzmann@ifado.de*

## Introduction

In a cocktail party scenario with at least two competing speakers, signal processing algorithms such as beamformers can benefit from information about the attended speaker. Auditory attention decoding (AAD) addresses this problem by predicting the attended speaker from recorded electro-encephalogram (EEG) signals. As shown in Figure 1, AAD decodes the speech envelope of the attended speaker $y_a(t)$ from the EEG signals and correlates it with the envelopes of the competing speakers. Following this general approach, several linear and non-linear decoding approaches have been developed [1]. While preceding work [2] showed, that the neural representation of speech indeed resembles the stimulus envelope, which motivates the decoding of speech envelopes from EEG signals, it is also known that the neural representation lags behind the stimulus onset. In our work, we therefore aim to increase the decoding accuracy of linear subject-specific decoders by compensating the individual neural latencies and by finding optimal filter orders for the decoder. To reduce the computational complexity of the decoder training, we also investigate the benefit of applying a dimensionality reduction method to the EEG signals, which takes the target speech envelope into account.

## Auditory attention decoding

Auditory attention decoding (AAD) refers to the decoding of the attended envelope from the EEG signal and comparing the estimated envelope with the envelopes of the speakers. To this end, the decoder can be implemented as an array of FIR filters, which filter the EEG signals to obtain the speech envelope of the attended speaker $y_a(t)$. The estimated envelope of the attended speaker can then be computed over the channel-wise convolution

$$y_a(t) = \sum_{c=1}^{C} \sum_{\kappa=0}^{K} h_c(\kappa) x_c(t - \kappa + L), \tag{1}$$

where $K$ is the filter order, $C$ is the number of channels, $h_c$ are the filter coefficients for channel $c$ and $x_c$ is the EEG signal for channel $c$. To compensate the neural latency between speech and the corresponding EEG response, we shift the EEG signal by $L$ samples. The objective is now to find the optimal filter coefficients and the best parameter set of filter order $K$ and lag $L$.

Ordinary least squares (OLS) regression is a well known technique to estimate optimal filter coefficients by minimizing the mean-squared error between prediction and observed variable. To determine the optimal filter coefficients, we re-write (1) in vector notation for a given speech segment $\mathbf{y}_i \in \mathbb{R}^N$

$$\mathbf{y}_i = \mathbf{X}_i \mathbf{h}_i, \tag{2}$$

where $\mathbf{h}_i \in \mathbb{R}^{CK}$ contains the corresponding filter coefficients for all EEG channels and $\mathbf{X}_i \in \mathbb{R}^{N \times CK}$ contains the EEG data. Using the pseudo-inverse, the OLS solution can be found by

$$\mathbf{h}_i = \left(\mathbf{X}_i^T \mathbf{X}_i\right)^{-1} \mathbf{X}_i^T \mathbf{y}_i. \tag{3}$$

The first term $\mathbf{X}_i^T \mathbf{X}_i$ is the correlation matrix of the EEG signals, while the second term $\mathbf{X}_i^T \mathbf{y}_i$ is the cross-correlation vector between EEG signals and speech envelope. We train the decoder by averaging the correlation matrices and the cross-correlation vectors over $I$ trials, where each trial consists of a speech and corresponding EEG segment.

Given the usually high number of partly correlated EEG channels, a direct application of the OLS method entails a high computational effort and possibly numerical instabilities because of ill-conditioned correlation matrices. Therefore, it is worthwhile to apply dimensionality reduction techniques to the EEG signals. One idea is to use principal component analysis (PCA), which transforms the EEG signals into a lower-dimensional space while retaining maximum variance components. The transformed representation can then be used for the OLS method instead of the EEG matrix $\mathbf{X}_i$. However, as PCA is agnostic about the target speech envelope, the principal components do not necessarily have high predictive power if only few components are retained. Another idea is to use the partial least squares (PLS) method, which transforms both, the prediction variable and the observable data into new spaces in which the transformed representations have maximal covariance. In our case, the prediction variable is the envelope of the attended speaker, while the observed data is the EEG data.

## Experiments

In our experiment we focus on subject-specific decoders and the subject-specific compensation of the neural latency. Therefore, we evaluate the decoding accuracy, which is the frequency how often the attended speaker is detected in all trials. Because the number of training trials is small, we perform a leave-one-out cross-validation and evaluate the decoder techniques (OLS, PLS) separately. As a baseline approach, we use a decoder with coupled filter order and decoder lag $K = L$ which takes all EEG samples from the current point in time to the

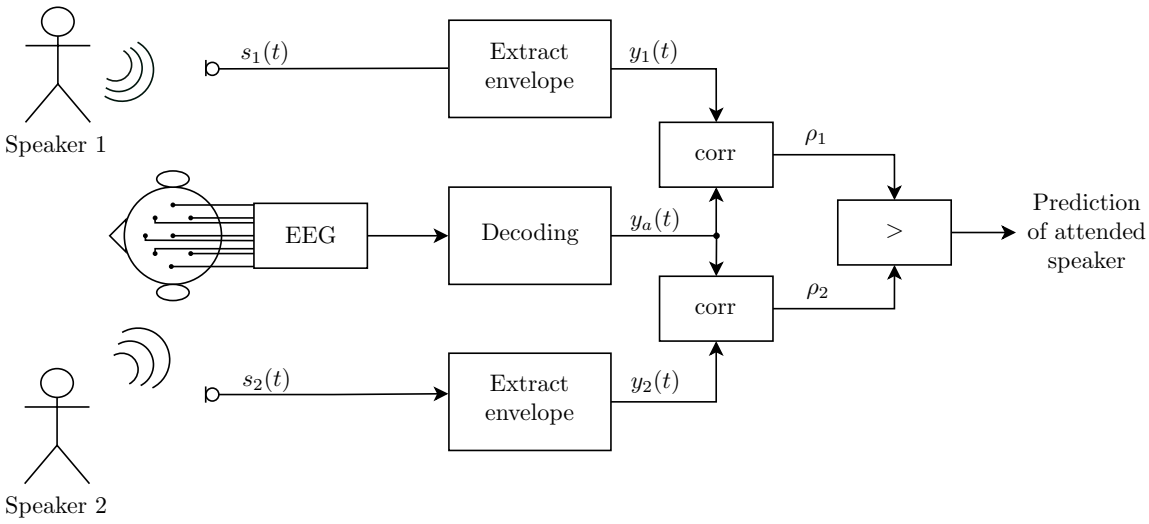


**Figure 1:** Structure of auditory attention decoding (AAD). AAD decodes the envelope of the attended speaker from the EEG signal and predicts the attended speaker by correlation of estimated envelope with the speech envelopes of the speakers.

lag $L$ into account. For the OLS method, we investigate the decoding accuracy as a function of lag $L$ and filter order $K$. Then, we perform an exhaustive search for the best parameter set (lag $L$ , filter order $K$) for each participant. For the PLS method, we additionally find the best number of components for each participant. As search space, we consider $L = 0 \dots 10$, $K = 0 \dots 9$, #PLS components $= 1 \dots 64$.

For our experiments, we use the dataset in [3], where the EEG signals of 16 participants (8 male and 8 female) have been recorded in a cocktail party scenario. The stimuli were played in two conditions *dichotic* and *HRTF*. For the dichotic condition, one speaker per ear was presented, whereas for the HRTF condition the speech signals were filtered with head-related impulse responses, so that one speaker was perceived from the left side and the other from the right side. We evaluate the conditions separately. A 64-channel EEG was recorded with a sampling rate of $f_s = 128$ Hz. The experiment was set up with two competing speakers, where the participants had the task to attend to one speaker and ignore the other.

We preprocessed the data (similar to [4, 5]), so that speech envelopes and estimated attended speech envelope can be compared. Therefore, we extract speech envelopes by applying a power-law $|\cdot|^{0.6}$ on the broad-band speech signals and then downsample the signals to 128 Hz. Now both signals have the same sampling rate. We filter both signals with a band-pass filter between $1 - 9$ Hz and decimate to 20 Hz, which is our target sampling rate. In the end, we split the signals into trials of 30 s.

## Results

We evaluate the decoding accuracy of the OLS decoder with respect to lag $L$ and filter order $K$ for the conditions dichotic and HRTF. Figure 2 shows the results with the baseline method (blue), which uses all the EEG data from the current time to a given lag $L$, a method with optimal filter order for each lag (red), a method with filter order $K = 0$ (green) and the proposed method,

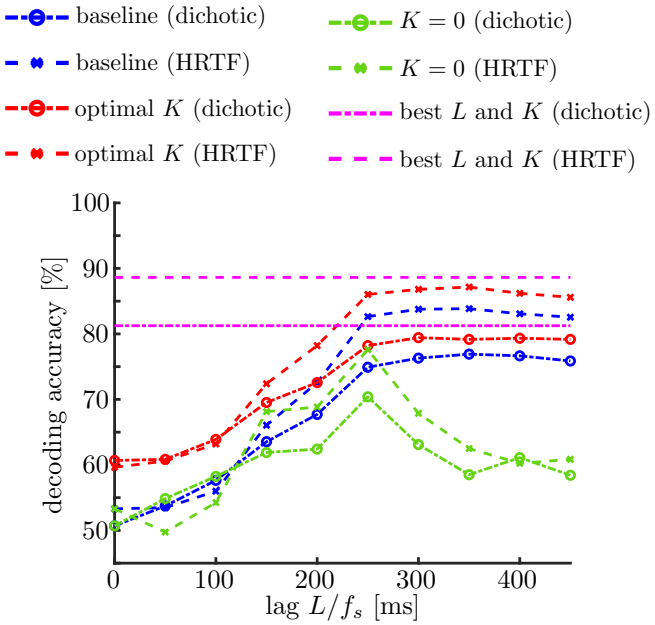


**Figure 2:** Decoding accuracy for the OLS decoder under the conditions dichotic (o) and HRTF (x) with different methods. The baseline method (blue) uses EEG data from the current time to a given lag $L$. The optimal $K$ method uses optimal filter order for each lag and subject (red). The method with filter order $K = 0$ uses only one sample per EEG channel (green). The proposed method (best $L$ and $K$) uses optimal lag and filter order for each subject (pink).

which uses optimal lag and filter order for each subject (pink). Comparing the conditions, we can see, that the HRTF condition outperforms the dichotic condition for $L/f_s > 150$ms. Focusing on the decoder with $K = 0$ where the maximum accuracy can only be achieved if EEG and stimulus envelopes are synchronized, we find the maximum decoding accuracy at 250ms which we define as the average neural latency. With this insight we analyze and compare the baseline method (blue) and the method with optimal filter order for each lag (red), where we can see an increase of performance up to 250ms. Using

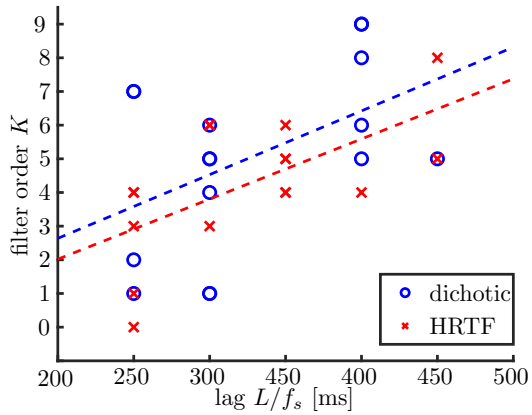


**Figure 3:** Distribution of best lag $L$ and filter order $K$ for the dichotic (blue circles) and HRTF (red crosses) conditions with corresponding regression lines.

a larger lag leads only to small changes in performance. Also, we can see, that the method with optimal filter order for each lag (red) outperforms the baseline method (blue). Finally the proposed method (pink line) with optimal lag and filter order for each subject, outperforms all other methods.

For the OLS decoder, we plot the optimal lag $L$ and corresponding filter order $K$ for each participant in Figure 3 where we evaluate the conditions dichotic and HRTF separately. The figure shows a linear relationship between lag and filter order. However, the baseline method ($K = L$) does not necessarily lead to the best combination of lag and filter order. Furthermore, we observe that the combination of lag and filter order always cover an average neural latency of 250ms. From the OLS result we conclude, that the subject-specific neural latency can be compensated by a good combination of lag $L$ and filter order $K$.

In Table 1 the mean decoding accuracy for OLS, PLS with all components ($\text{PLS}_{\text{all}}$) and PLS with the best number of components ($\text{PLS}_{\text{best}}$) are presented. Comparing the performance between the PLS methods, $\text{PLS}_{\text{best}}$ outperforms $\text{PLS}_{\text{all}}$. We further analyze the best number of PLS components across all participants in Figure 4. The figure shows the best number of components which are substantially lower (13 for dichotic, 12 for HRTF) compared to the full rank representation (64 components). This result confirms the hypothesis that dimension reduction does not only reduce the computational complexity but also increases the decoding accuracy.

Table 1 also compares the proposed methods statistically by a right-sided t-test with the baseline methods. We can see, that all proposed methods outperform the baseline methods highly significantly ($p < 0.001$). As already observed for the OLS method, we can also see here a better performance for the HRTF condition than for the dichotic condition. The PLS method with all components ($\text{PLS}_{\text{all}}$) does not necessarily outperform the OLS method, but finding the best number of components increases the accuracy. On average, $\text{PLS}_{\text{best}}$ outperforms OLS for the dichotic condition (increase of 2.65% in performance). For the HRTF condition, the differences between $\text{PLS}_{\text{best}}$ and OLS are small (increase of 0.78% in performance).

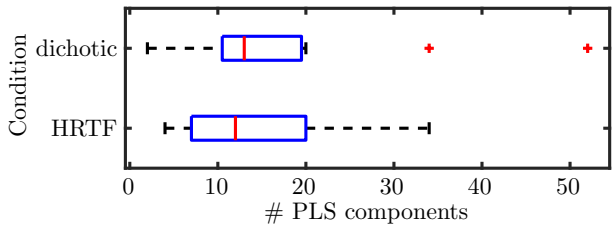


**Figure 4:** Best number of PLS components across all subjects for the conditions dichotic (median 13) and HRTF (median 12).

**Table 1:** Results for OLS, PLS with all components ($\text{PLS}_{\text{all}}$), PLS with the best number of components ($\text{PLS}_{\text{best}}$) for the proposed method and the baseline method. The conditions are evaluated separately and a right-sided t-test was performed. All proposed methods outperform the baseline method highly significantly ($p < 0.001$). The baseline method uses EEG data from the current time to a given lag $L$, whereas the proposed method always uses the best lag and filter order for each subject.

| | | accuracy [%] | | |
|---|---|---|---|---|
| | method | proposed | baseline | p-value |
| dichotic | OLS | 81.25 | 76.91 | $1.65\text{x}10^{-6}$ |
| | $\text{PLS}_{\text{all}}$ | 79.43 | 72.22 | $2.62\text{x}10^{-6}$ |
| | $\text{PLS}_{\text{best}}$ | 84.90 | 80.73 | $2.74\text{x}10^{-6}$ |
| HRTF | OLS | 88.63 | 83.85 | $1.75\text{x}10^{-6}$ |
| | $\text{PLS}_{\text{all}}$ | 85.94 | 74.48 | $1.21\text{x}10^{-5}$ |
| | $\text{PLS}_{\text{best}}$ | 89.41 | 83.77 | $4.44\text{x}10^{-5}$ |

## Conclusions

In this paper we have proposed a method, which optimizes the lag and filter order of auditory attention decoders (AAD) to compensate the subject-specific neural latency. The results show a significant increase of the overall decoding accuracy compared with generic settings of the lag and filter order. Hence, a subject-specific compensation of the neural latency and choice of optimal individual decoder filter orders appear to be crucial for an improved decoder performance. AAD applications, which rely on deep learning methods, could therefore also benefit from such a subject-specific adjustment. For this purpose appropriate estimation approaches have to be developed. Regarding the PLS method, our investigation shows that reducing the number of components leads to an improved decoding performance. In particular, the optimal number of components turns out to be much lower than the maximum possible number of components, which can possibly be explained by the target-oriented dimensionality reduction performed by PLS. Future AAD algorithms could therefore benefit from a joint optimization of the subject-specific latency and a PLS-based decoding scheme.

## Acknowledgement
This work has been funded in parts by Deutsche Forschungsgemeinschaft (DFG) within SPP Audictive, project number 444777670, and Collaborative Research Center SFB 823, subproject B3.

## References
[1] Geirnaert, S. et al.: Electroencephalography-Based Auditory Attention Decoding: Toward Neurosteered Hearing Devices. IEEE Signal Processing Magazine 38 (2021), 89-102

[2] O'Sullivan, J. et al.: Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. Cerebral Cortex 25 (2015), 1697-1706

[3] Das, N. and Francart, T. and Bertrand, A.: Auditory Attention Detection Dataset KULeuven (2020), URL: `https://doi.org/10.5281/zenodo.3997352`

[4] Biesmans, W. et al.: Auditory-Inspired Speech Envelope Extraction Methods for Improved EEG-Based Auditory Attention Detection in a Cocktail Party Scenario. IEEE Transactions on Neural Systems and Rehabilitation Engineering 25 (2017), 402-412

[5] Das, N. et al.: The effect of head-related filtering and ear-specific decoding bias on auditory attention detection. Journal of Neural Engineering 13 (2016), 056014 (9pp)