# Predicting speech understanding from EEG recordings: effect of attention

Jonas Vanthornhout[1] Lien Decruy[1] Jan Wouters[1] Jonathan Z. Simon[2,3,4] Tom Francart[1]

[1]KU Leuven - University of Leuven, Department of Neurosciences, ExpORL [2]Department of Electrical and Computer Engineering, University of Maryland, College Park, USA [3]Department of Biology, University of Maryland, College Park, USA [4]Institute for Systems Research, University of Maryland, College Park, USA

KU LEUVEN

## 1. Introduction

### Objective correlate of speech understanding

- Speech envelope is a primary cue for speech understanding (Shannon et al., 1995)
- Cortical neural activity tracks the neural envelope of running speech (Peelle and Davis, 2012)
- Reconstruction of speech envelope from cortical activity is possible (Ding and Simon, 2011)
- Quality of this reconstruction correlates with behaviourally measured speech understanding (Vanthornhout et al., 2016)

### Research question

**Question 1:** Does the reconstruction quality of the speech envelope depend on how much attention the subject paid to the stimulus? Hypothesis: $attention(subject) \downarrow \Rightarrow SNR(EEG) \downarrow \Rightarrow reconstruction\ quality \downarrow$

**Question 2:** Can we remove/reduce the effect of attention by choosing optimal signal processing parameters?

## 2. Methods

### Participants

3 experiments:

1. 33 young normal hearing subjects, aged 21-29 years
2. 5 young normal hearing subjects, aged 22-29 years
3. 8 young normal hearing subjects, aged 19-23 years

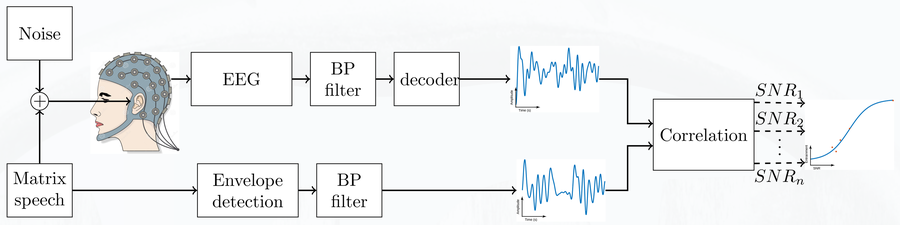
### EEG

- BioSemi system with 64 electrodes
- Flemish story to train linear decoder: 15 minutes
- Listening to Flemish Matrix speech sentences at $>5$ SNRs
- 2-4 repetitions

### 3 EEG conditions

1. Maximal attention: anwering questions about the speech stimuli
2. Watching an animated movie
3. Playing tetris

### Signal processing




$$decoder = (RR^T)^{-1}(RS^T)$$

$R$ time-lagged matrix of the neural data

$S$ stimulus envelope of story

Correlation the reconstruction quality is measured by calculating the Spearman correlation between the actual and reconstructed envelope.

BP filter a delta band filter (0.5 Hz to 4 Hz).

$$\hat{s}(t) = \sum_n \sum_\tau decoder(n,\tau) R(t+\tau, n)$$

$\hat{s}$ reconstructed envelope

$t$ time ranging from 0 to $T$

$n$ recording electrodes ranging from 1 to $N$

$\tau$ post-stimulus samples used to reconstruct the envelope: integration window

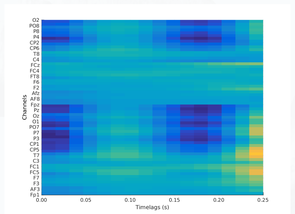



Figure: An example of a decoder. A decoder is a mixing matrix of channels and time lags.

Integration window of the decoder can influence the effect of attention (Puvvada and Simon, 2015; O'Sullivan et al., 2014)

- 0 ms to 75 ms: less modulated by attention
- >75 ms: modulated by attention
- 170 ms to 250 ms: very prominent effect of attention

We investigated the effect of different integration windows between 0 ms and 250 ms on the reduction of effect of attention. Reducing the effect of attention is important in an audiological application as the subject's motivation should not influence the results.

## 3. Results

### Experiment 1: effect of integration window (baseline)

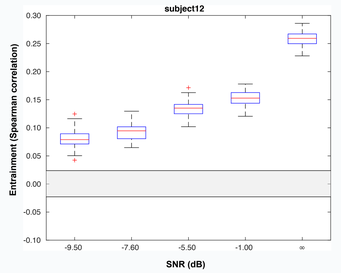



Figure: Example of *stimulus SNR* $\uparrow \Rightarrow$ *reconstruction quality* $\uparrow$. Correlations outside the grey band are significant. The used speech material was the Flemish Matrix, we used a grand-average decoder with an integration window ranging from 0-75 ms and a filterband from 0.5 Hz until 4 Hz.

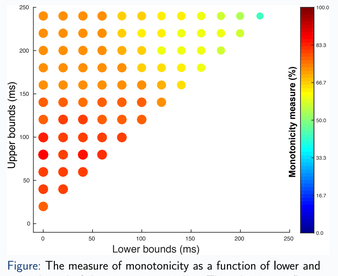



Figure: The measure of monotonicity as a function of lower and upper bound of the integration window. The subject-specific decoder performs best with a temporal integration window from 0-75 ms.

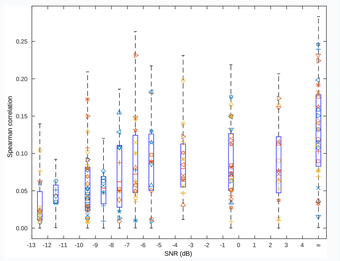



Figure: The reconstruction quality at different SNRs for 33 subjects. Each marker is one subject using a grand average decoder with an integration window from 0 ms until 75 ms. This figure shows that the correlation increases as a function of stimulus SNR (Spearman's $\rho = 0.85, p < 0.001$).

The measure of monotonicity checks if the reconstruction quality at a higher SNR is higher than the reconstruction quality at a lower SNR. If the reconstruction quality increase monotonically with SNR (e.g. the left hand figure) the percentage correct will be 100%. We found that in most subjects an integration window from 0-75 ms resulted in the highest percentage correct (middle figure).

### Experiment 2: effect of watching a movie

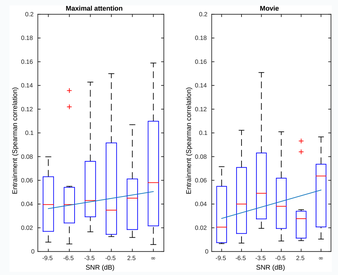



Figure: Integration window: 0-250 ms

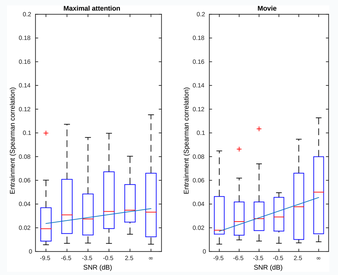



Figure: Integration window: 0-75 ms

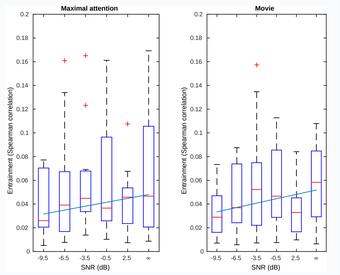



Figure: Integration window: 75-250 ms

The entrainment of the stimulus at multiple SNRs with and without watching a movie. The integration window 0-75 ms shows in both conditions a quasi monotonical trend suggesting the effect of attention was less prominent compared to the other integration windows.

Visually, we see that the increase of reconstruction quality is more monotonic compared to the other integration windows, this is supported by the correlation between the reconstruction quality and the SNR for the movie condition (Spearman $\rho = 1.0$ for 0-75 ms is significantly higher than $\rho = 0.5$ for 0-250 ms and $\rho = 0.6$ for 75-250 ms using Zou's confidence interval.)

### Experiment 3: effect of playing Tetris

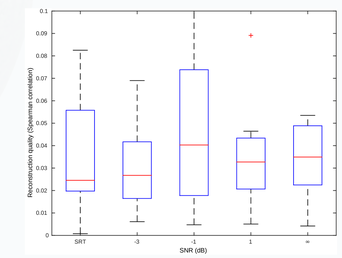



Figure: Integration window: 0-75 ms. Watching Tetris

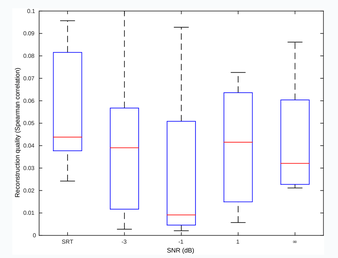



Figure: Integration window: 0-75 ms. Playing Tetris

- SRT of the subjects was obtained using a behavioural Flemish Matrix test, SRTs were around $-8$ dB SNR.
- Reconstruction quality increases quasi-monotonically when subjects were attentively listening to speech sentences while watching a Tetris game (Spearman $\rho = 0.7$).
- When playing a Tetris game, the reconstruction quality did not increase with SNR (Spearman $\rho = -0.5$).
- These results do not correspond with the results from the previous conditions. However, playing Tetris introduced many motor artefacts, something which was not present in the previous conditions.

## 4. Conclusion

1. We found that entrainment of the speech envelope increases with stimulus SNR when choosing an integration window of 0-75 ms. Our hypothesis is that this results from having less effect of attention which can influence the level of entrainment (Puvvada and Simon, 2015; O'Sullivan et al., 2014).
2. When subjects were watching a movie while listening to the speech stimuli, we found monotonically increasing entrainment with stimulus SNR in the 0-75 ms integration window. Other integration windows did not show a monotonic increase.
3. When subjects were playing Tetris while listening to the speech stimuli, we did not find this monotonic increase. We found the monotonic increase when the same subjects attentively listened to the speech stimuli. While this can be the result of failing to remove the effect of attention, it can also be due to motor artefacts.

**Research question 1:** When the attention of the subject was diverted to a movie or a Tetris game, we found a decrease of reconstruction quality at some SNRs.

**Research question 2:** By choosing an integration window from 0 75 ms we were able to reduce the effect of attention in the movie condition but not in the Tetris condition.

## References

Nai Ding and Jonathan Z. Simon. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, 2011.

James A. O'Sullivan, Alan J. Power, Nima Mesgarani, Siddharth Rajaram, John J. Foxe, Barbara G. Shinn-Cunningham, Malcolm Slaney, Shihab A. Shamma, and Edmund C. Lalor. Attentional selection in a cocktail party environment can be decoded from single-trial eeg. *Cerebral Cortex*, 2014.

J. E. Peelle and M. H. Davis. Neural Oscillations Carry Speech Rhythm through to Comprehension. *Front Psychol*, 2012.

Krishna Puvvada and Jonathan Z. Simon. Time-constrained neural decoding from multiple auditory cortical areas. *Association for Research in Otolaryngology, 39th MidWinter meeting (Baltimore, Maryland)*, 2015.

R V Shannon, F G Zeng, V Kamath, J Wygonski, and M Ekelid. Speech recognition with primarily temporal cues. *Science*, 1995.

Jonas Vanthornhout, Lien Decruy, Jan Wouters, Jonathan Z. Simon, and Tom Francart. Speech understanding can be predicted from brain recordings. *In preparation*, 2016.

## Acknowledgements

Financial support for this project is provided by the KU Leuven Special Research Fund under grant OT/14/119. Research funded by a PhD grant of the Research Foundation Flanders (FWO). This project has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No 637424), and from the YouReCa Junior Mobility Programme under grant JUMO/15/034 to Jonas Vanthornhout. The authors thank Lise Goris and Eline Verschueren for helping collecting data.

## QR code

A digital version of this poster can be found on `https://exporl.med.kuleuven.be/web/index.php/Public:IHCON2016` or by scanning the QR code.