

# EEG-based Decoding of Selective Visual Attention in Superimposed Videos: Supplementary Material

Yuanyuan Yao, Wout De Swaef, Simon Geirnaert, and Alexander Bertrand, *Senior Member, IEEE*

This supplementary material accompanies the paper titled “EEG-based Decoding of Selective Visual Attention in Superimposed Videos” by Yao et al [1].

## I. PERFORMANCE IN SVAD AND MM USING EYE-RELATED DATA MODALITIES

Unlike in EEG-based decoding (Section III-D of the main text), the SVAD task yields comparable or even higher accuracies than the MM task when using eye-related data modalities, as shown in Fig. 1a. Comparable results are expected if the unattended object is not encoded in the eye-related data modalities, as “unattended” would equate to “unobserved”. However, the higher accuracies in SVAD are counterintuitive and warrant further investigation, although they do not influence the results and conclusions of the paper.

One possible explanation is that limited data leads to insufficient sampling. To illustrate this, we compare the distributions of the correlations of the *ObjFlow* i) between randomly sampled pairs of attended and unattended video segments that were presented together in the SVAD task, and ii) between randomly sampled pairs of attended and mismatch segments in the MM task. With sufficient sampling, we expect that both distributions have a similar shape, peaking around zero. However, as observed in Fig. 1b, a negative bias is present in the correlation between attended and unattended 30-second segments, which is reduced when using 5-second segments. Note that the number of sampled pairs is approximately 4600 in both tasks and for both segment lengths. However, the sampled 30-second segments are more likely to overlap with each other, while the 5-second segments tend to be more independent. This suggests that insufficient sampling due to limited data is the cause of the bias. The bias is not observed in the correlation between attended and mismatch segments, likely because the sampling of mismatch segments allows to introduce much more variability: the mismatch segment can be randomly drawn from the set of all unobserved segments, whereas the unattended segment is always paired one-to-one with the attended segment. This negative bias may make the unattended segments easier to identify than mismatch segments, contributing to higher accuracies in SVAD.

One way to correct for the bias without expanding the current dataset is to use partial correlation instead (Section II-E.2 of the main text). We control for the influence of the unattended object’s feature when calculating the correlation between the data and the feature of the attended object, and vice versa. The decoder then makes decision based on the partial correlations. However, the results are similar to the

original, which suggests that the previously mentioned bias might not be the primary factor behind the unexpected results.

Another potential explanation relates to participants’ behavior. To better ignore the distractor, participants may suppress eye movements toward the unattended object and fixate their gaze when the unattended object has large motion and becomes particularly distracting. This inhibition of distractors in eye-related data modalities could result in slightly better performance for SVAD compared to MM. However, this hypothesis remains speculative and is difficult to verify with the current setup.

## REFERENCES

- [1] Yuanyuan Yao, Wout De Swaef, Simon Geirnaert, and Alexander Bertrand. EEG-based decoding of selective visual attention in superimposed videos. 2024. URL <https://arxiv.org/abs/2409.12562>.

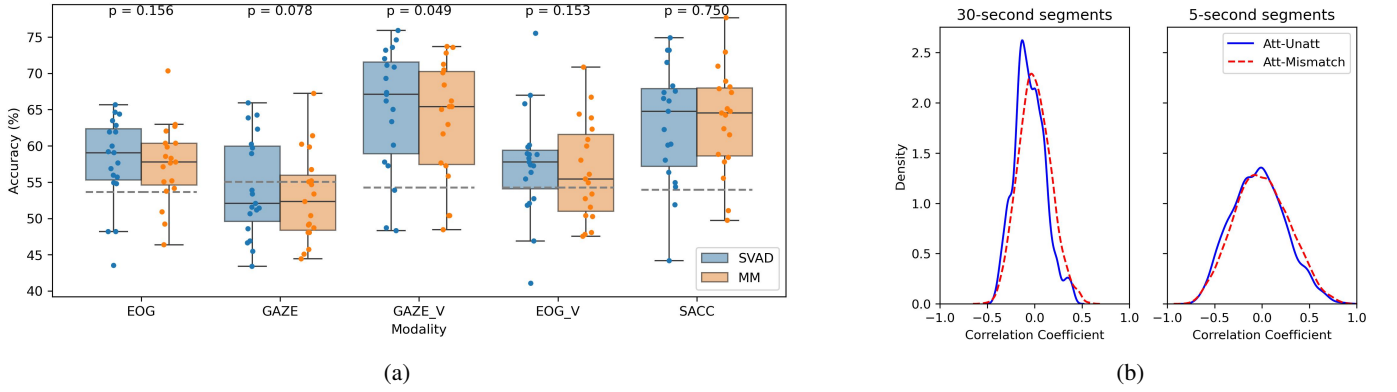


Fig. 1: (a) Accuracies of the SVAD and MM tasks for eye-related modalities. The models are trained and tested on the superimposed-object dataset and the test segments are 30s long. The dots denote the individual (per-subject) accuracies, and the boxes show the median and the interquartile range. Wilcoxon signed-rank tests are performed to determine if the accuracies of SVAD tasks are significantly higher than that of MM tasks using EEG. (b) Distributions of correlations between attended and unattended video features in the SVAD task (in blue) versus attended and mismatch video features in the MM task (in red) when testing with 30-second (left) and 5-second (right) segments.