EEG exploring the time-window feature [May 9]

This text continues the results of May 2nd. See the header below.

What part of the ERP is most relative to classification?

Hypothesis. The classification quality depends on 1) the time the measurement started and 2) the length of the time window (measured segment).

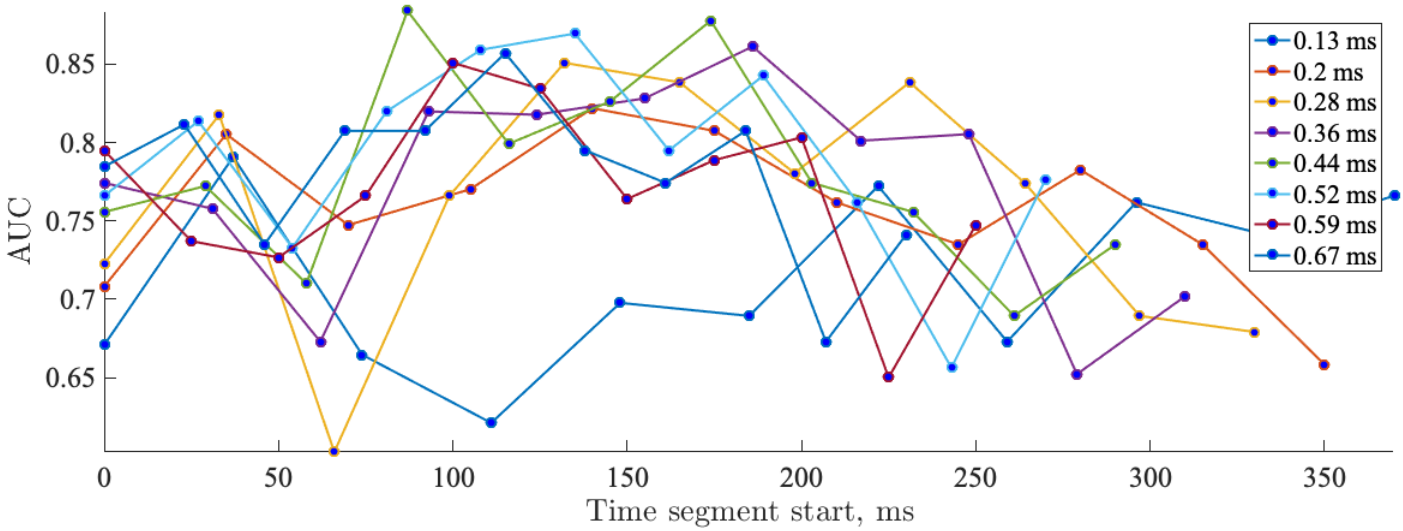

Figure 1. Classification accuracy for various onset of the time segment given the length of the time segment. Zero corresponds to the Event onset. The segment lengths are 0.130 … 0.670 ms, and the step is 50 ms.

Conclusion. We can significantly drop the time window (0.52 – 0.36 ms) without losing AUC. The blue window with the low AUC is 0.13 ms wide. The window can start after 60 ms. The yellow window 0.28 ms wide is decent after 100 ms.

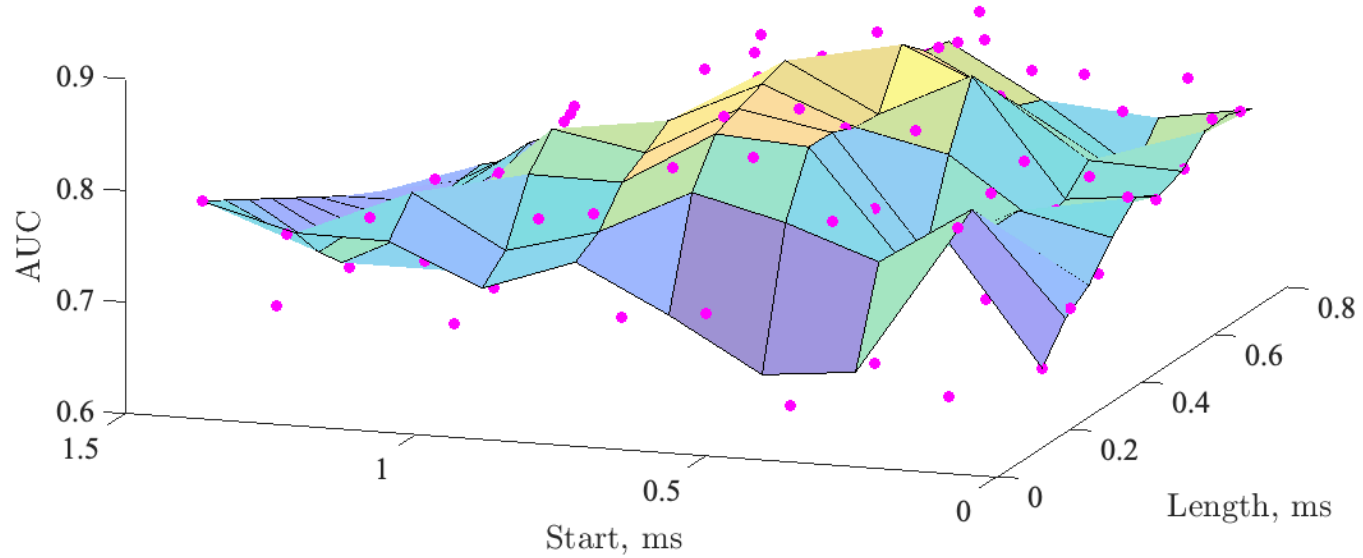

Figure 2. The same as Figure 1 in 3d format. The Classification accuracy for various lengths of the time segment. The length varies from 0.125 ms to 0.725 ms.

Header

News of Friday 6th: All useful preprocessing and feature exptaction are already programmed in NeurpPype. And it makes sense to exclude programmed algorithms to avoid doubling. Also, after discussion, we agreed that all the tested adequate models deliver approximately the same accuracy. So as it was already concluded, the focus is the scientific interpretation.

The waiting list for access to utilities

Utility	Purpose
AWS data bucket UMN	To access 1) the data made on my request, 2) the augmented dataset with 30+ users
AWS virtual machine	To run Neuropypelines
PyCharm on the AWS VM	To run NeuroPype API
AWS / WPI VM resources	To run computational experiments
AWS data bucket WPI	To keep intermediate results
Matlab VM	To get FIFx viewers (I have 2017 home edition)
Matlab Toolboxes	Preprocessing
Access to UH	On weekends and evenings
Mail cell	A large quality screen, etc.

Experiment condition 1 (to refer)

The Event-Related Potential (ERP) starts in the stimulus onset of and lasts 1200 ms. The data include EEG recorded from 128 channels BioSemi headcap at 256 Hz sampling rate.

- The quality criterion is accuracy and AUC on the cross-validated data.
- Models: 1) Gaussian Process Classification, 2) Bayesian Naive Classification, 3) Logistic Regression. The features are generated by the First Principal Components of the extended covariance matrix. The covariance is between pairs of electrodes' time series.

A variant of Figures 1 and 2.

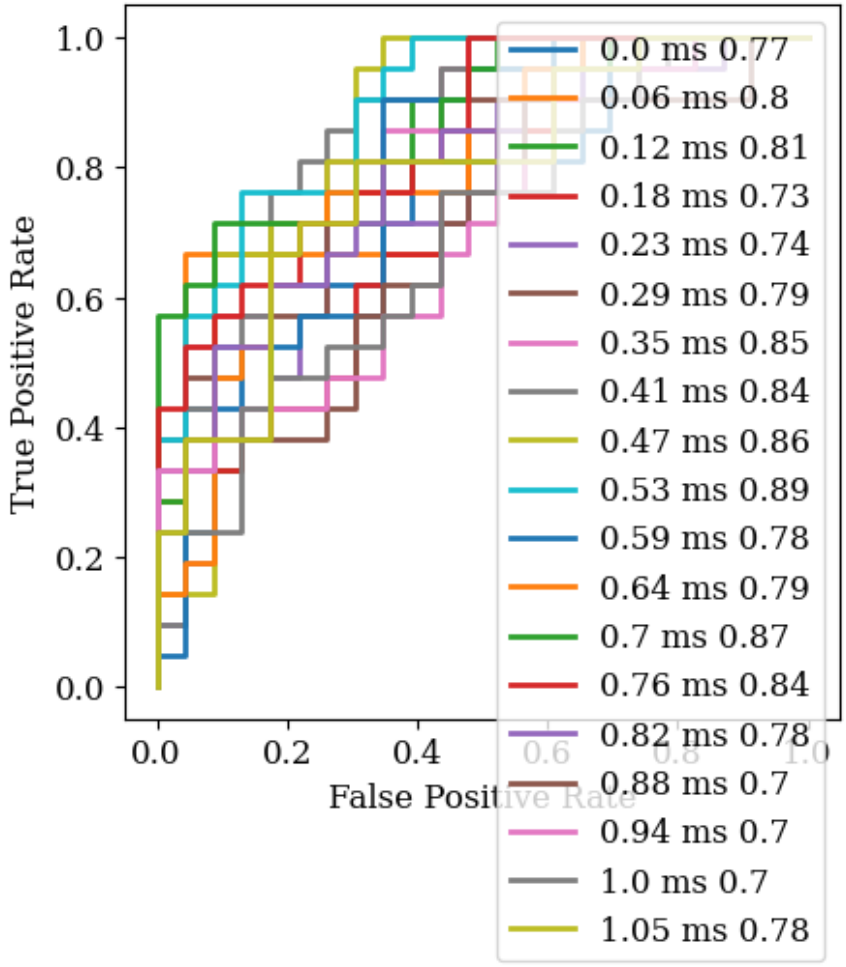

Figure 3. ROC and AUC for all time segments of fixed length 500 ms. Each segment starts as the legend shows.

Comment about slides 2023–05, slide 8

Caveat: This dataset has relatively few trials per subject since the study analyzed several other experimental conditions not of interest here. The data set contains about 80 events per participant x 13 participants vs 128ch x 450 time points comprising the EEG response. Each user has a unique pattern of brain signals, which creates obstacles to merging them into one common dataset. We aim to augment these data and mitigate the data heterogeneity issue (e.g., using standardization) to aid model performance.

(This does not read so well: The goal is to augment and ameliorate this data to make classification models concerning the user's unconscious perception only.)
(Removed: It contains events with noisy redundant signals that could be classified concerning the user's responses. - since that's well known about EEG data)

VerbMem Data status: published here - https://journals.sagepub.com/doi/full/10.1177/21677026211025018

Note on imported data, elimination records

The raw files collect a little bit more user events.

User	UMN	RAW
1034	**108**	**109**
1037	97	97
1045	90	90
1158	105	105
1363	**106**	**110**
1368	83	83
1385	75	75
2038	91	91
6639	**100**	**101**
7974	**90**	**91**
7977	110	110
7980	107	107
1327	59	59