Decoding EEG Data Using Machine Learning

Dr. Firat Soylu
The University of Alabama
http://fsoylu.people.ua.edu

Sabancı University, CS Seminar

June 16, 2023

Motivation

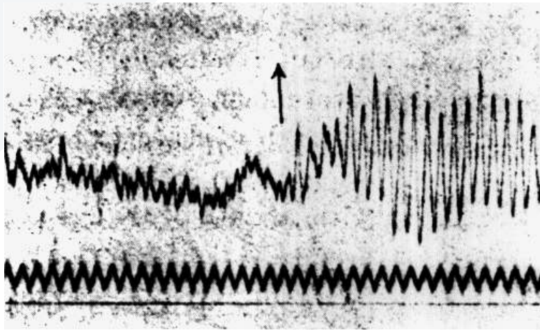
- Improve analysis of EEG data using machine learning (ML) methods.
- Develop heuristics for analyzing neuroscience data with ML methods.

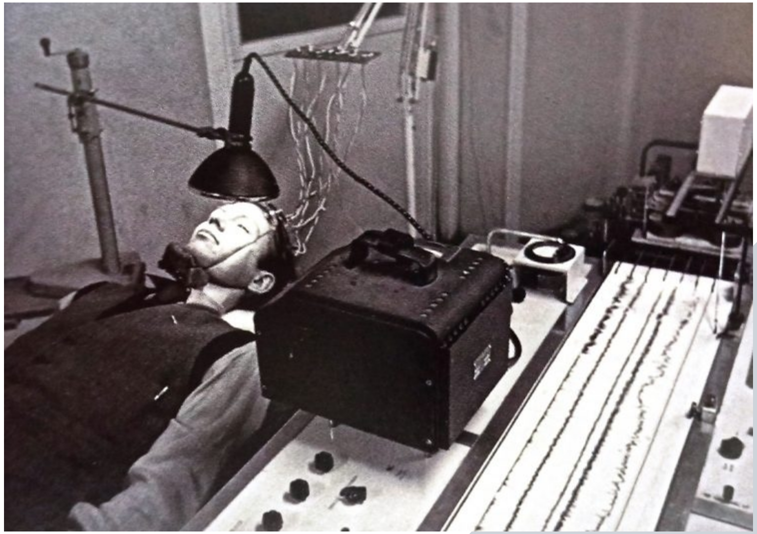
Outline

- EEG (Electroencephalogram)
- Use of EEG in cognitive neuroscience
- Traditional EEG data analysis methods
- Machine learning with EEG (and other neuro) data
- A path for future

Invention of EEG

- Hans Berger (1929)
  - Also discovered Alpha waves
- Initially met with skepticism





Electroencephalogram

- Used for studying sensory, attentional & cognitive processes
- Cheap
- Non-invasive
- Easy to use with children

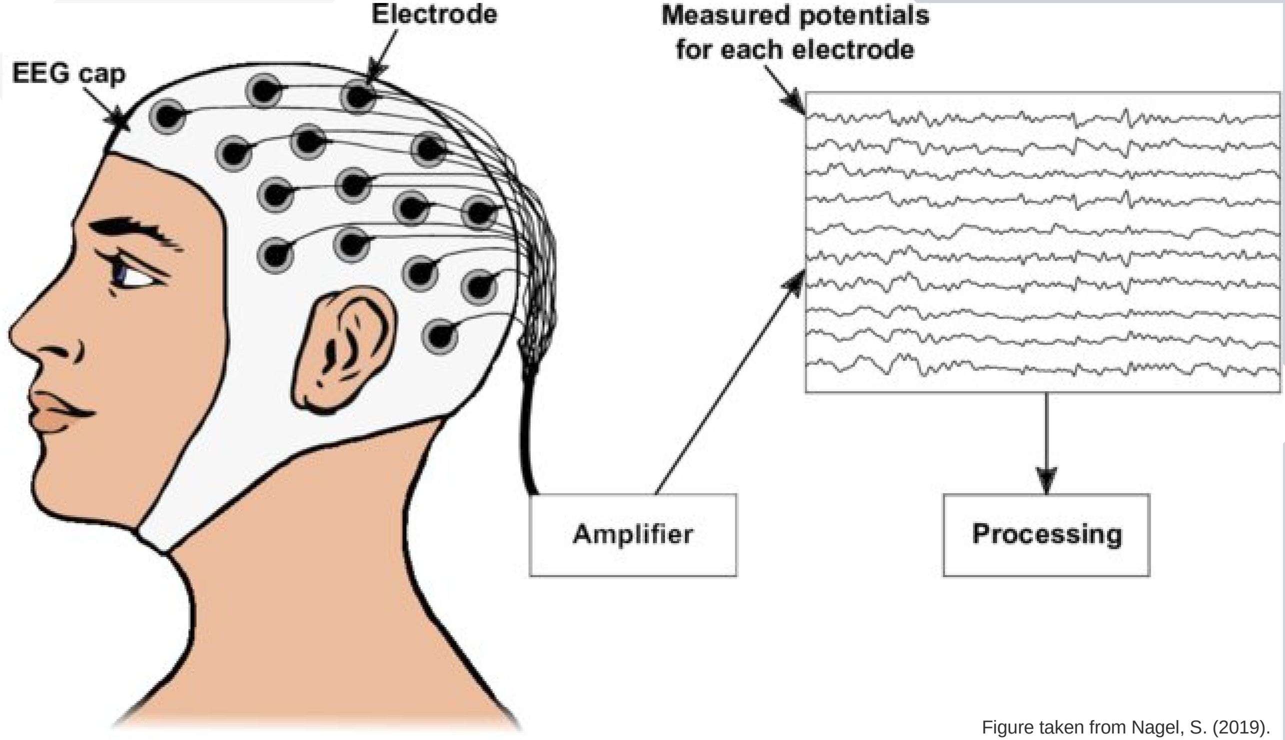
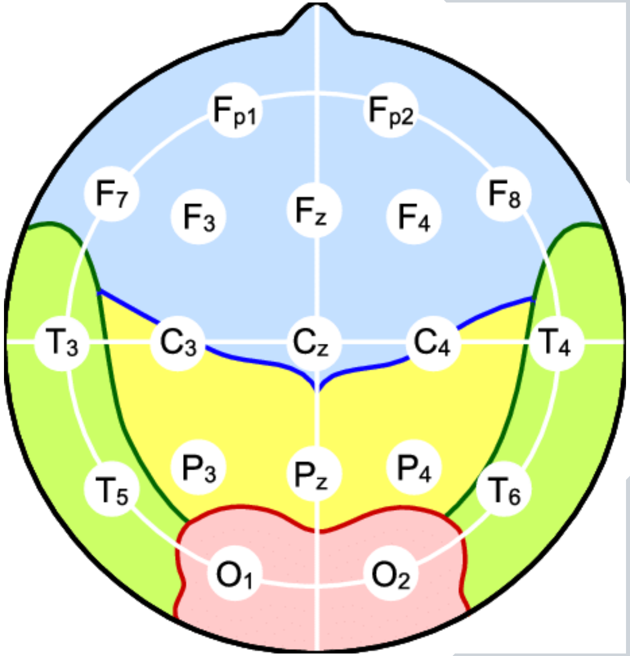


Figure taken from Nagel, S. (2019).

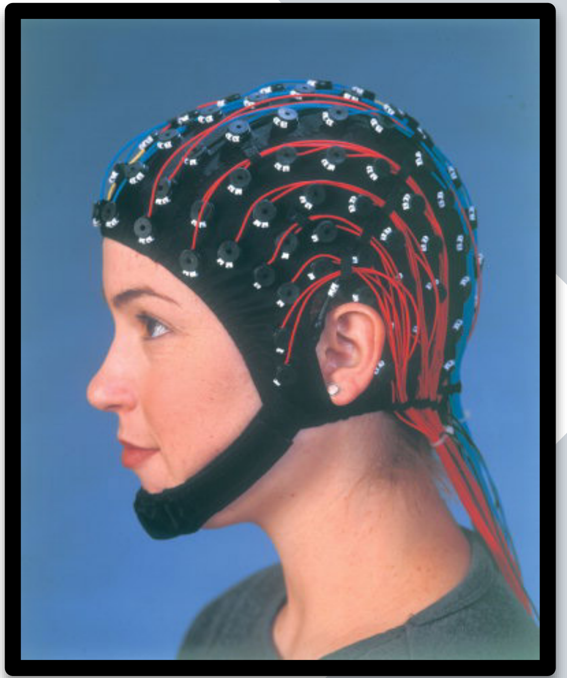
What does EEG measure?

- Voltage in each electrode relative to a reference electrode
- Signals are weak and have to be amplified (x10K-50K)
- Sampling rate ranges from 500 – 2000 Hz
- Low/high pass filters to filter noise

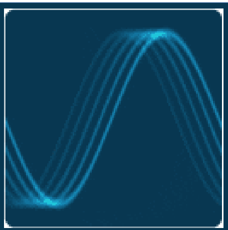


Principles of EEG

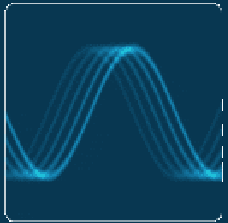
- Measures electrical activity in the (mostly) cortex
- Activity oscillates at different frequencies
- Measured in hertz (Hz) cycles/sec
- Most activity is between 0-100 Hz



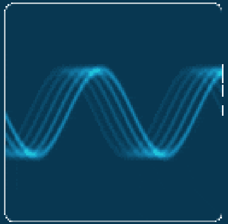
Frequency Bands



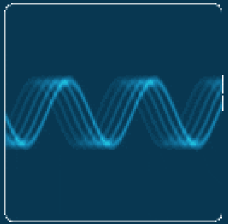
Delta (1 – 3 Hz)



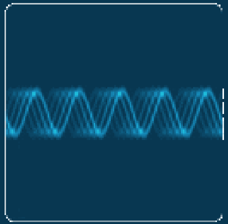
Theta (4 – 7 Hz)



Alpha (8 – 12 Hz)

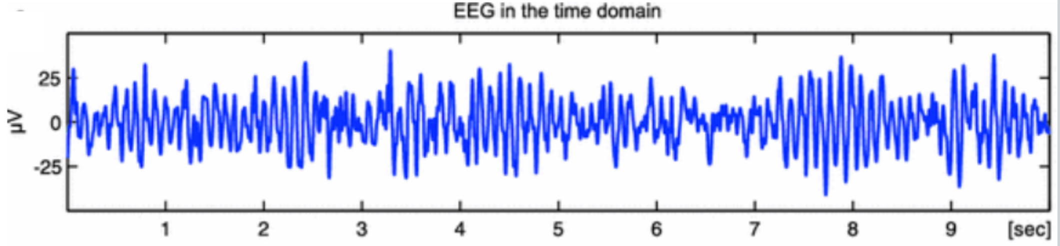


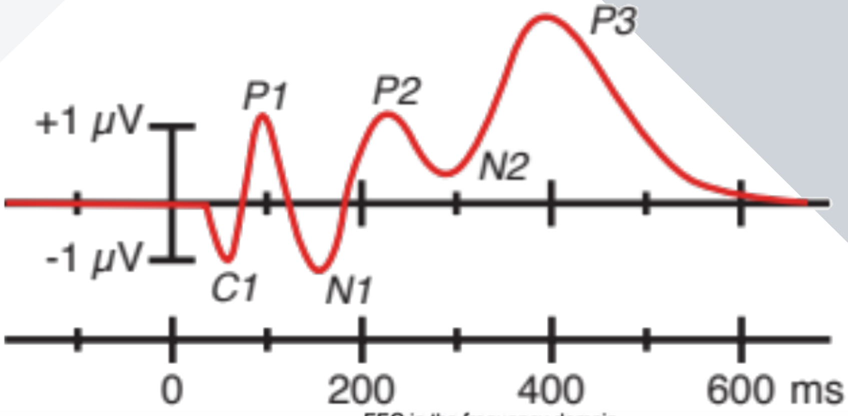
Beta (13 – 35 Hz)

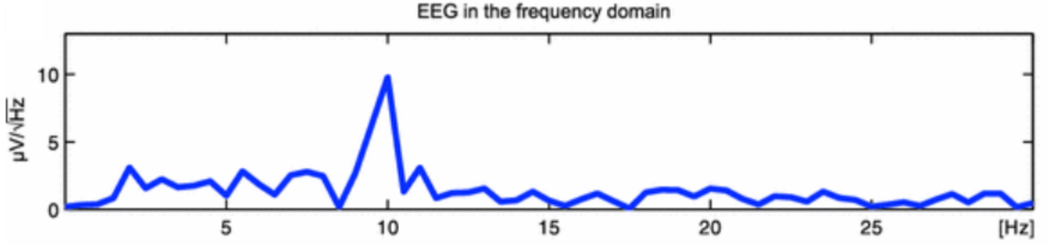


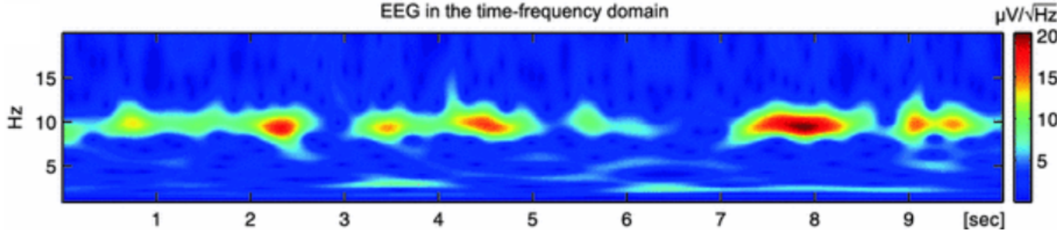
Gamma (36 – 100 Hz)

- Event-related potentials
- Frequency analysis
- Time-frequency analysis



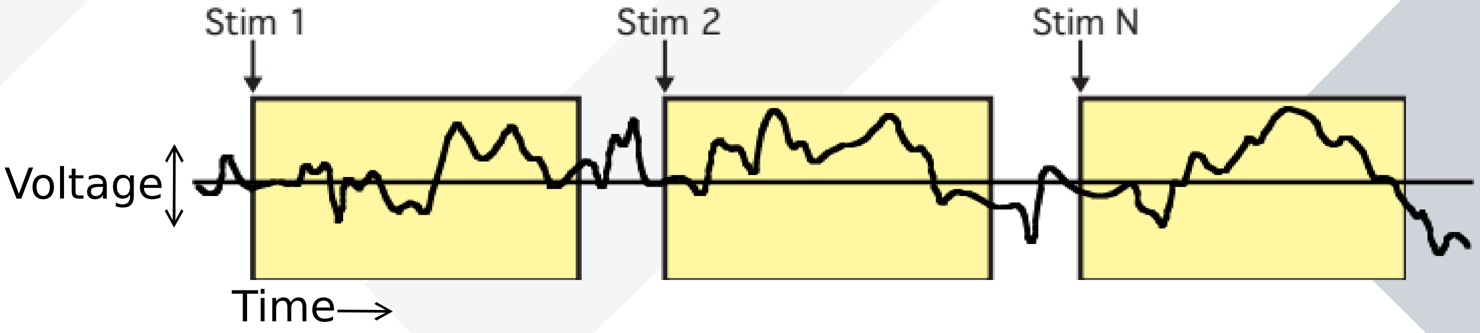






https://www.jonsprouse.com/courses/eeg-methods/time.frequency.pdf

ERPs



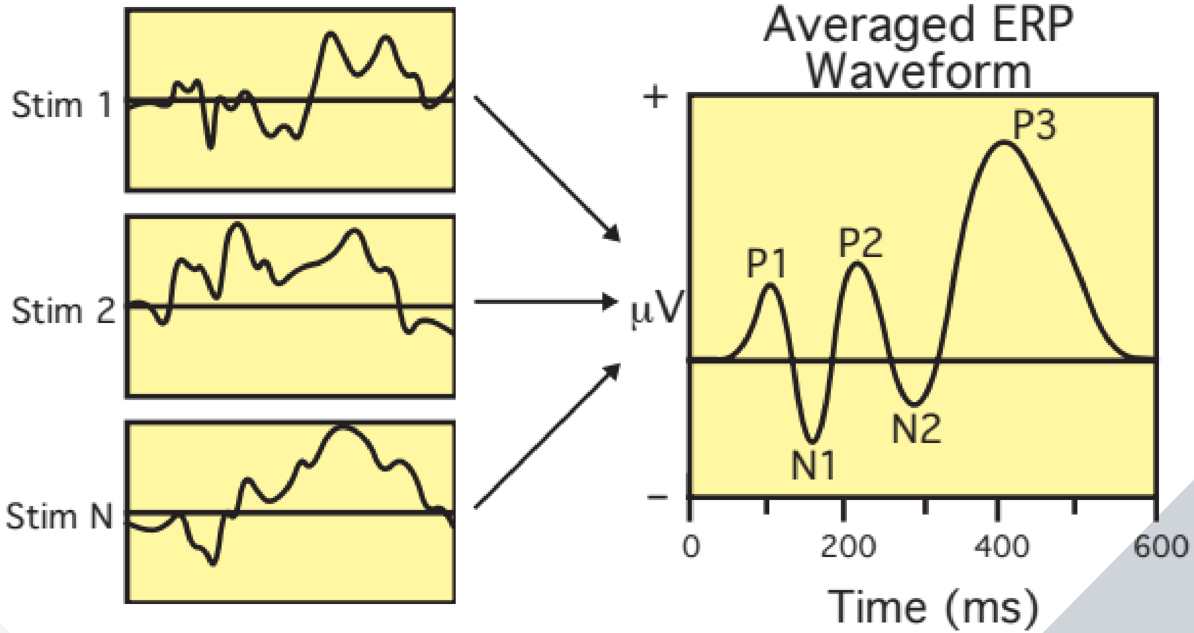
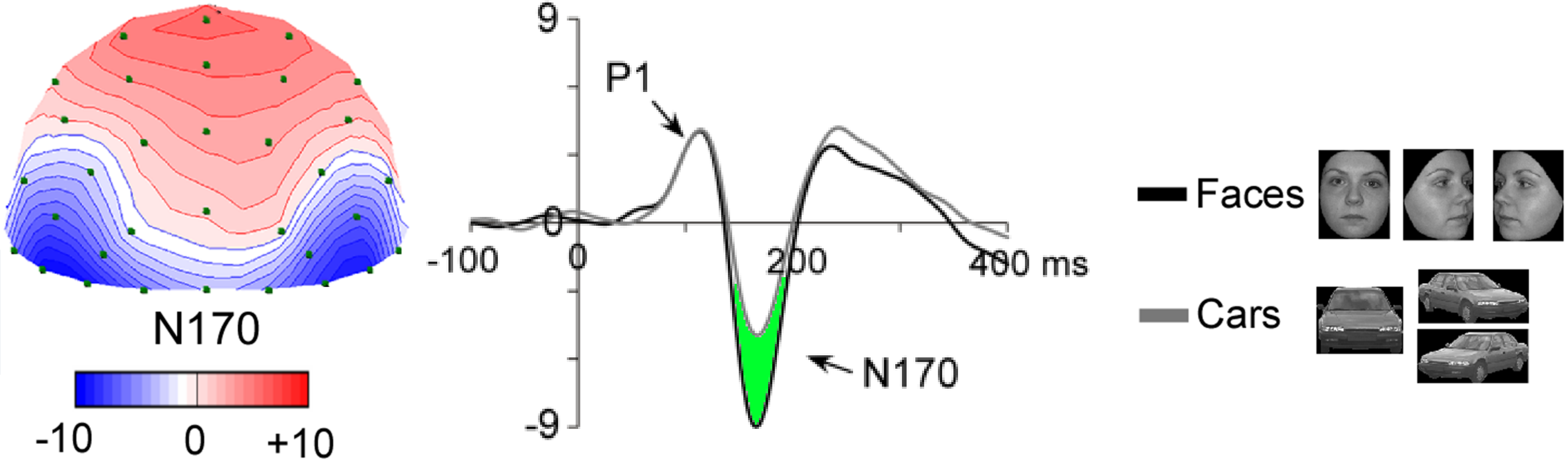


Figure: Courtesy of Steve Luck, ERP Bootcamp

Example: N170



Rossion & Jacques (2009)

N400

- Violation of a semantic expectation

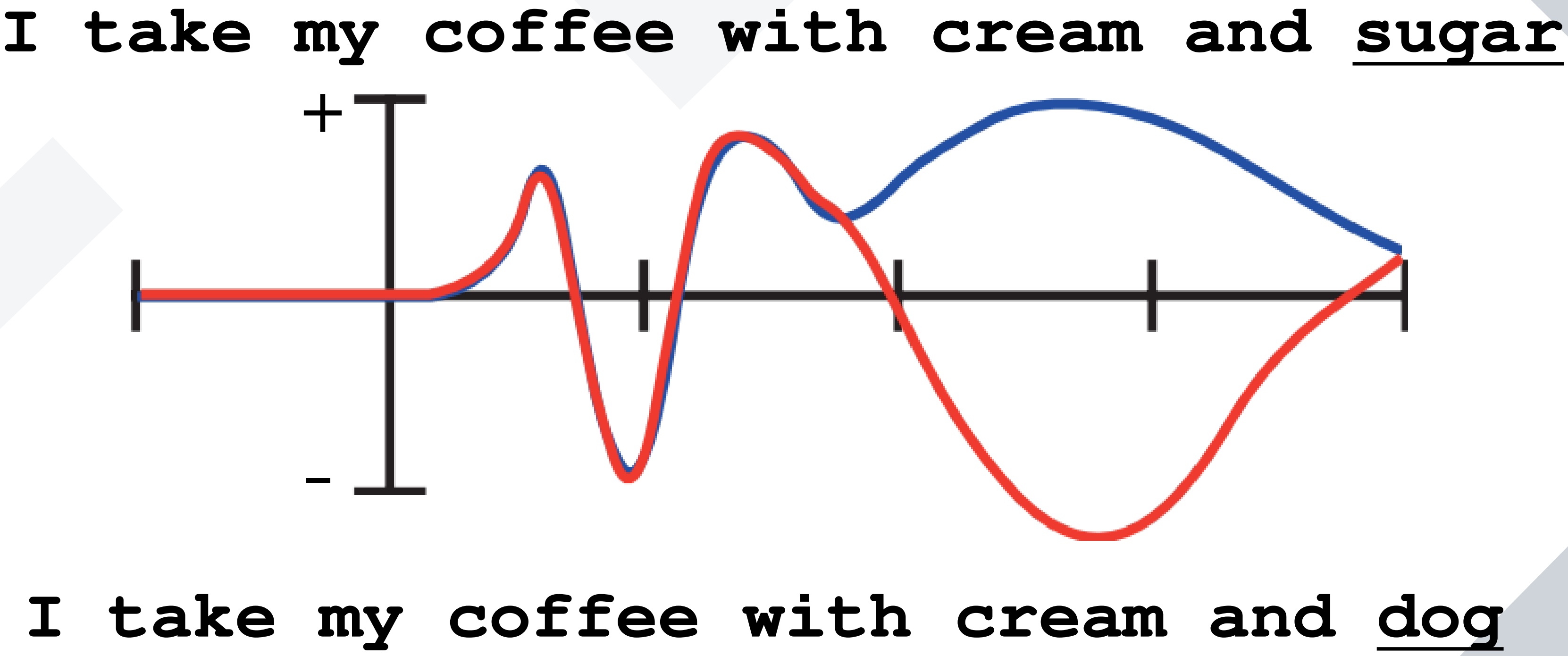
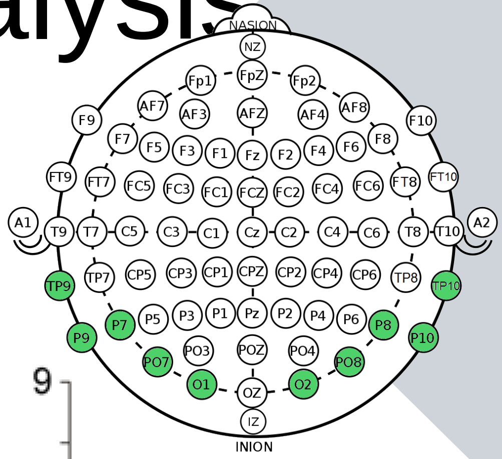
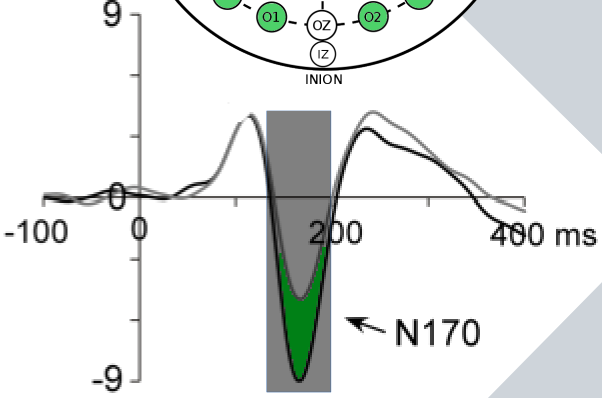


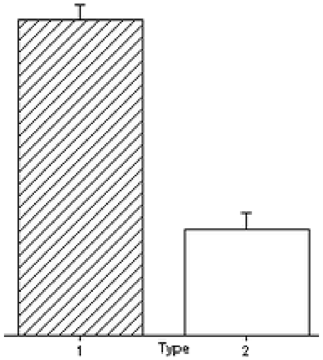
Figure: Courtesy of Steve Luck, ERP Bootcamp

Traditional EEG/ERP Analysis

- Determine relevant electrode sites
- Determine relevant time intervals
- Extract measurements
- Conduct inferential stats



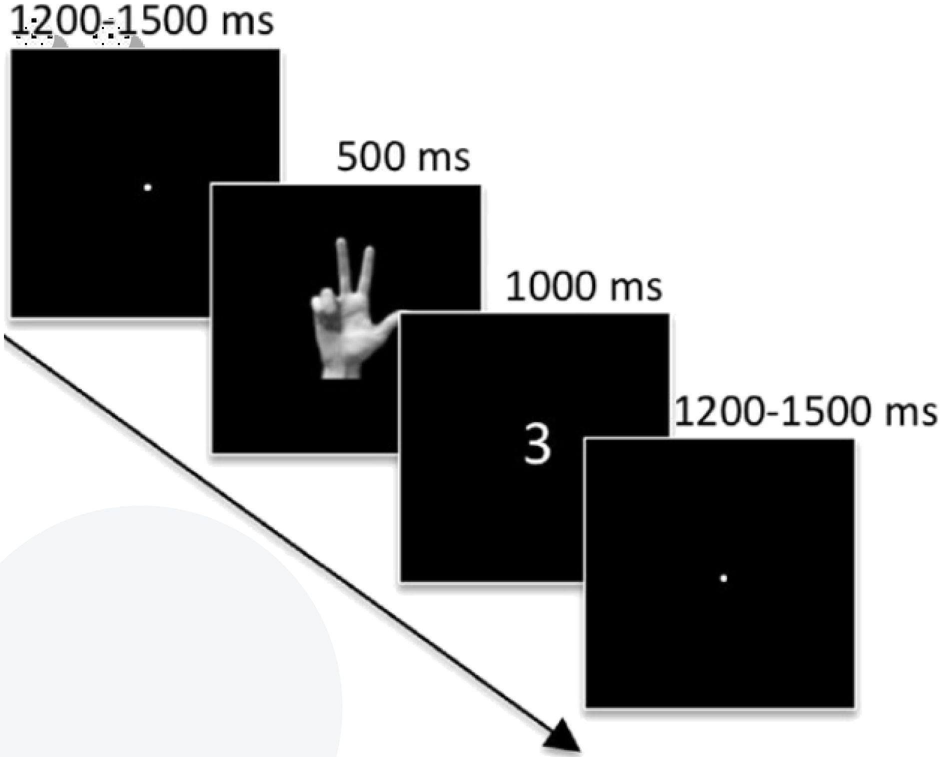




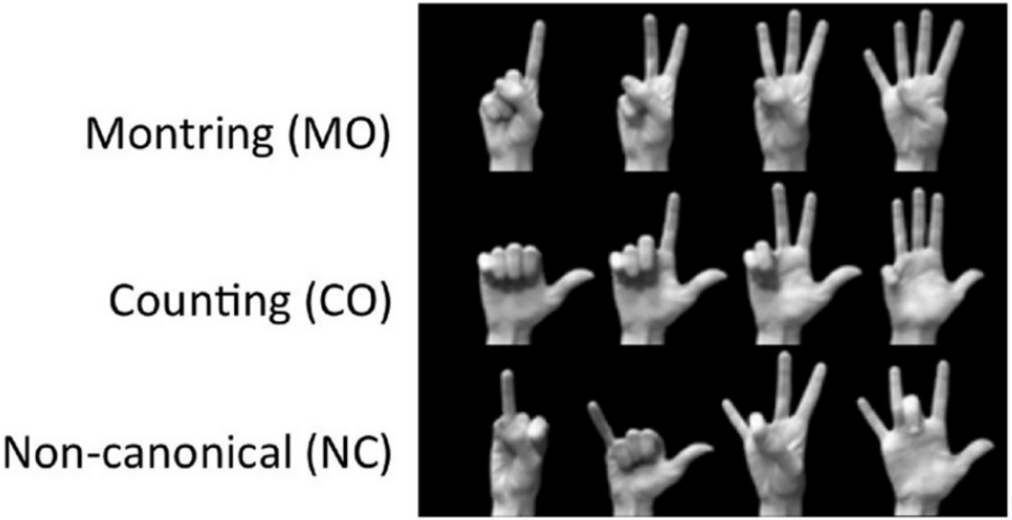
Enter Machine Learning ...

- Fewer a priori assumptions about the spatial location and time intervals for effects
- Allows testing effects distributed across different electrode sites and time intervals
- Good for conducting research in new domains

Case Study



- 38 adult participants (13 female)



Soylu, F., Rivera, B.*, Anchan, M.*, & Shannon, N. (2019). ERP differences in processing canonical and noncanonical finger-numeral configurations. Neuroscience Letters, 705, 74–79. https://doi.org/10.1016/j.neulet.2019.04.032

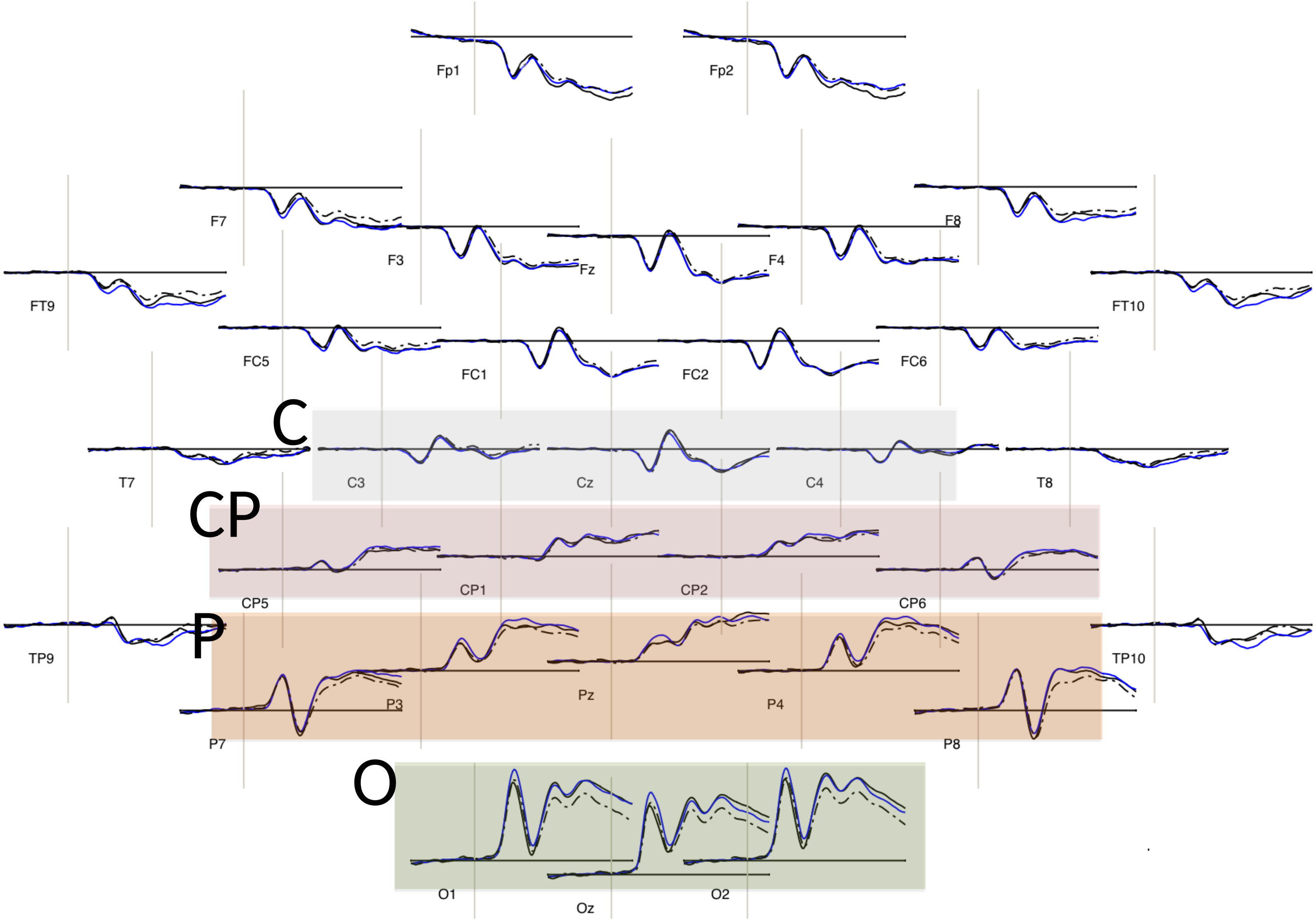
Soylu, F. (2019). Public dataset: ERP differences in processing canonical and noncanonical finger-numeral configurations, *Harvard Dataverse*.

Research Question:

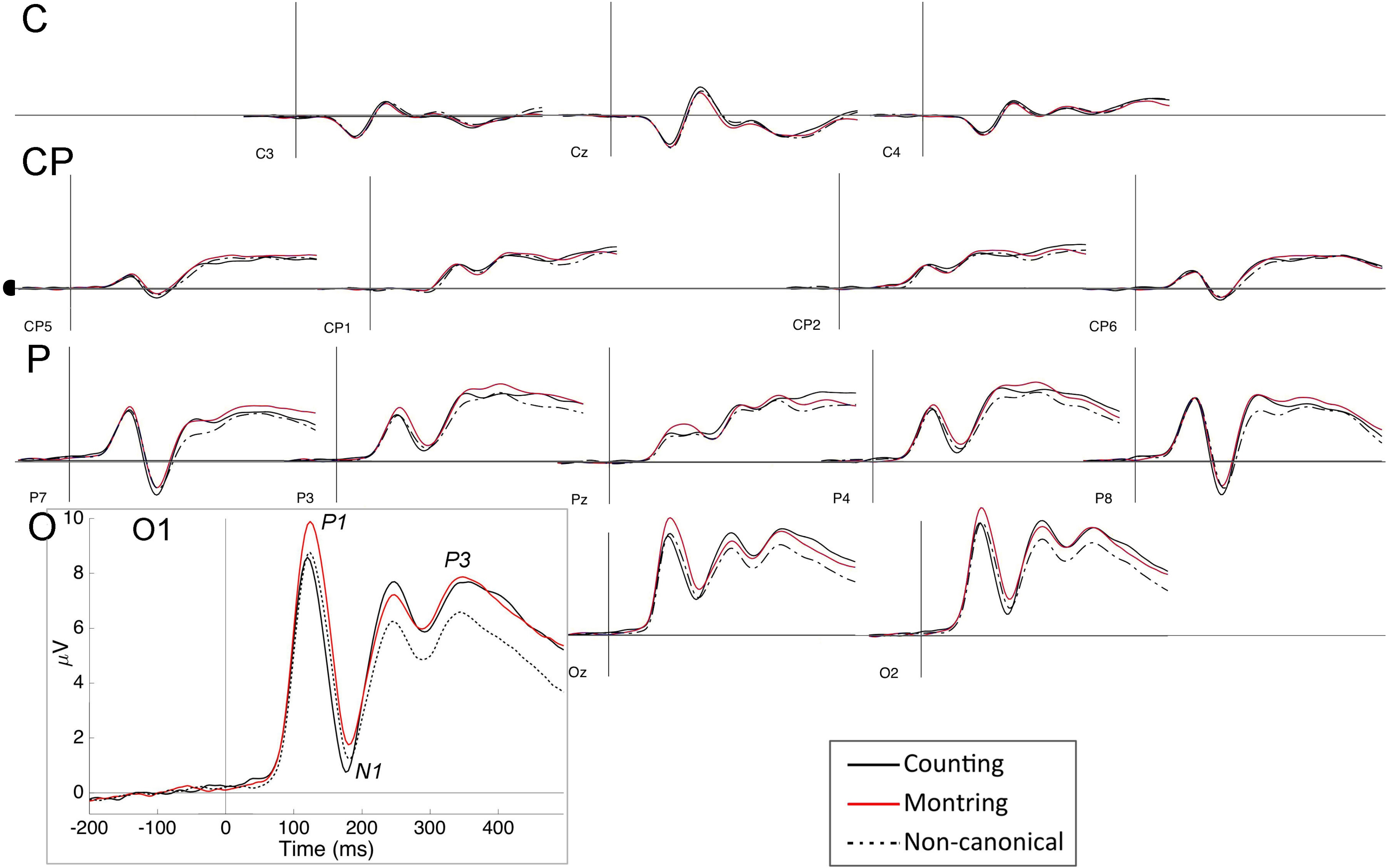
Do previous experiences with montring and counting lead to higher automaticity?


DELL

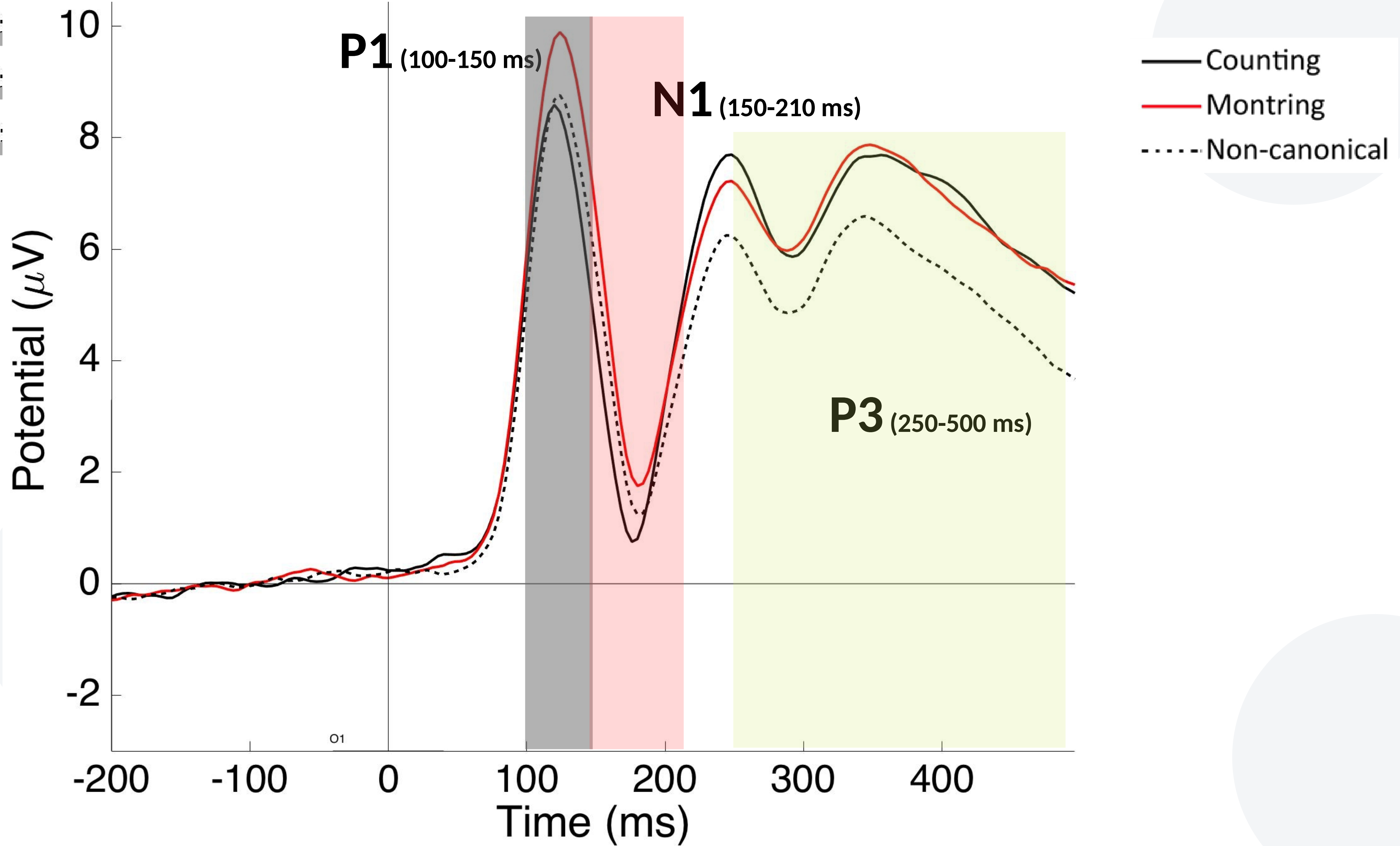
- Cl
  –



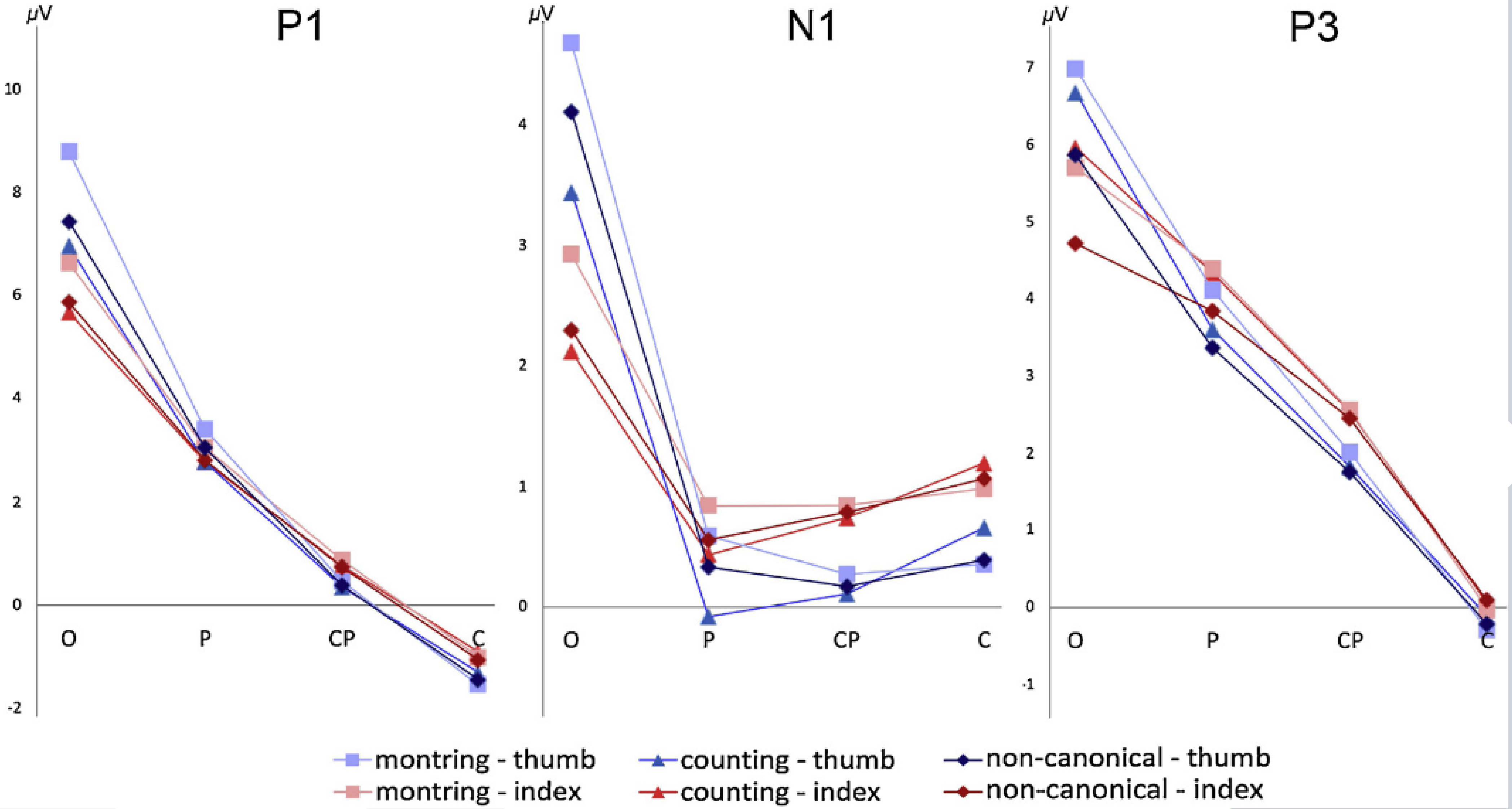
Fp1
Fp2
F7
F3
Fz
F4
F8
FT9
FT10
FC5
FC1
FC2
FC6
C
T7
C3
Cz
C4
T8
CP
CP5
CP1
CP2
CP6
P
TP9
TP10
P3
Pz
P4
P7
P8
O
O1
Oz
O2



C
C3
Cz
C4
CP
CP5
CP1
CP2
CP6
P
P7
P3
Pz
P4
P8
O
O1
P1
P3
N1
Oz
O2
10
8
6
4
2
0
μV
-200
-100
0
100
200
300
400
Time (ms)
Counting
Montring
Non-canonical



P1 (100-150 ms)
N1 (150-210 ms)
P3 (250-500 ms)
Counting
Montring
Non-canonical
Potential (μV)
Time (ms)
O1
10
8
6
4
2
0
-2
-200
-100
0
100
200
300
400



μV
P1
10
8
6
4
2
0
-2
O
P
CP
C
μV
N1
4
3
2
1
0
O
P
CP
C
μV
P3
7
6
5
4
3
2
1
0
-1
O
P
CP
C
montring - thumb
counting - thumb
non-canonical - thumb
montring - index
counting - index
non-canonical - index

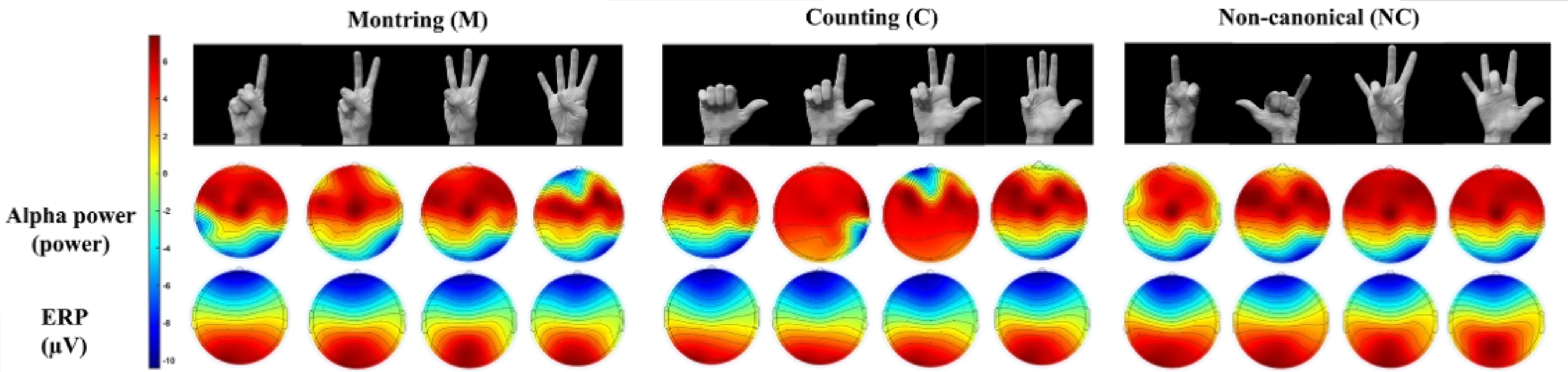
Results

- Higher behavioral performance for montring
- Higher P1/N1 positivity for montring
  - Feature-based attention system, focusing on features matching with a template.
- Similar P3 for montring and counting
  - Memory allocation as opposed to use of counting or subitizing

“The real voyage of discovery consists not in seeking new landscapes, but in having new eyes.”

-Marcel Proust

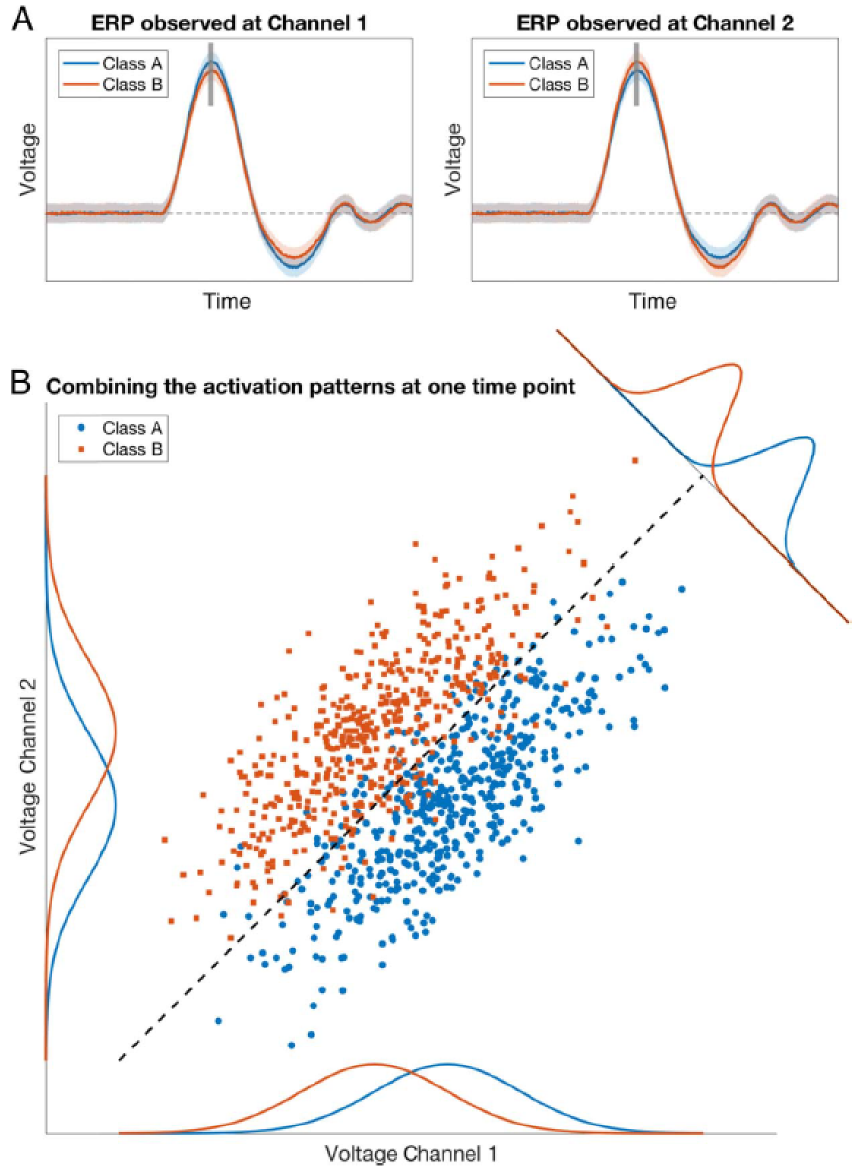
Decoding with SVM



Salehzadeh, R.*, Rivera, B., Man, K., Jalili, N., & Soylu, F. (2023). EEG Decoding of Finger Numeral Configurations With Machine Learning. *Journal of Numerical Cognition*, 9(1), 206-221. https://doi.org/10.5964/jnc.10441

What is decoding?

- Scalp distributions (SD) are varied across tasks & subjects
- Decoding with SD can allow prediction of task-related processes



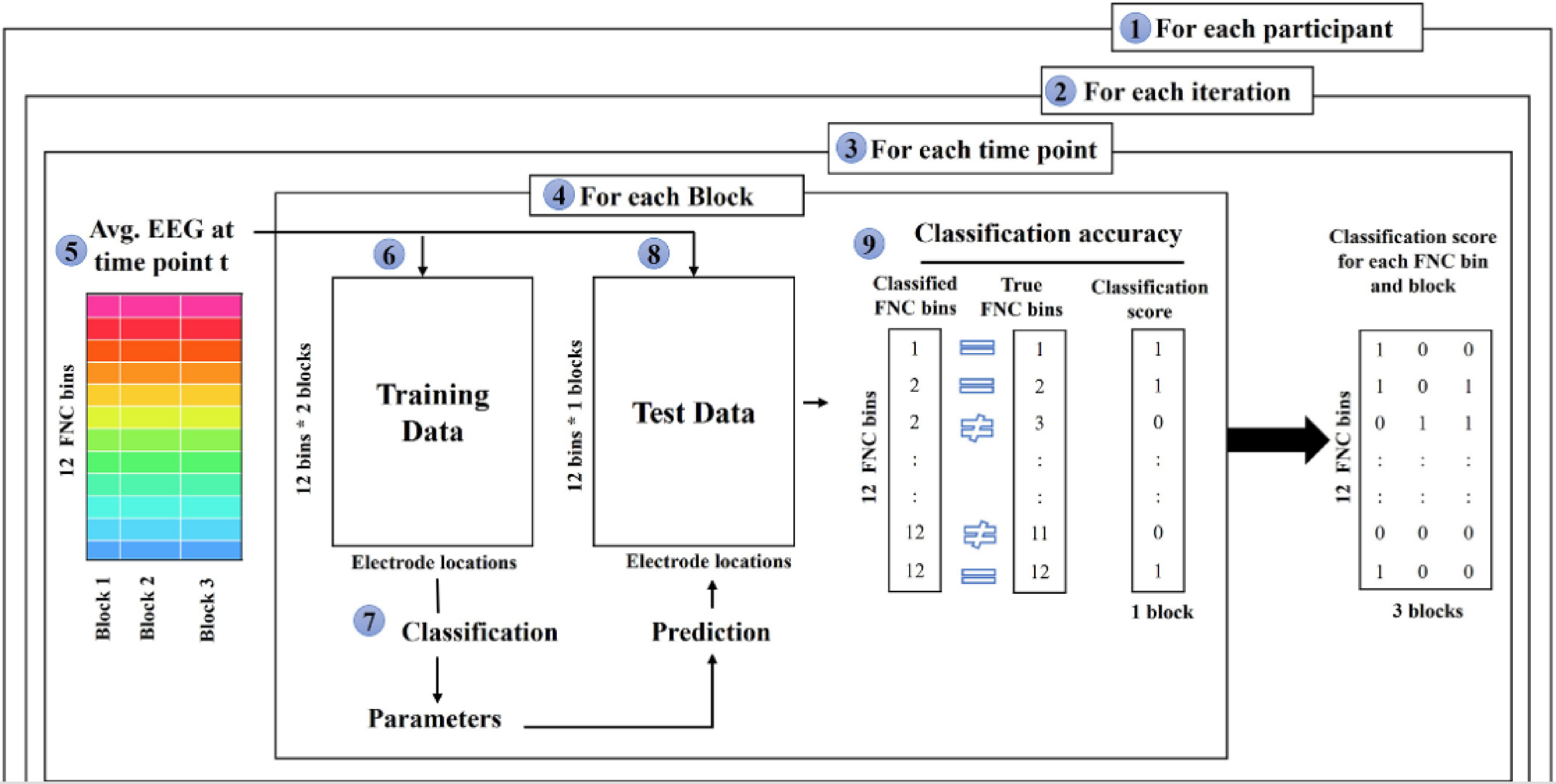
Grootswagers, Wardle, and Carlson (2017)

Decoding in Neuroscience

- Lower accuracy than engineering applications (e.g., brain-machine interfaces)
- The goal is to test research hypotheses
- Data SNR is crucial
- Preprocessing steps might differ from traditional analysis

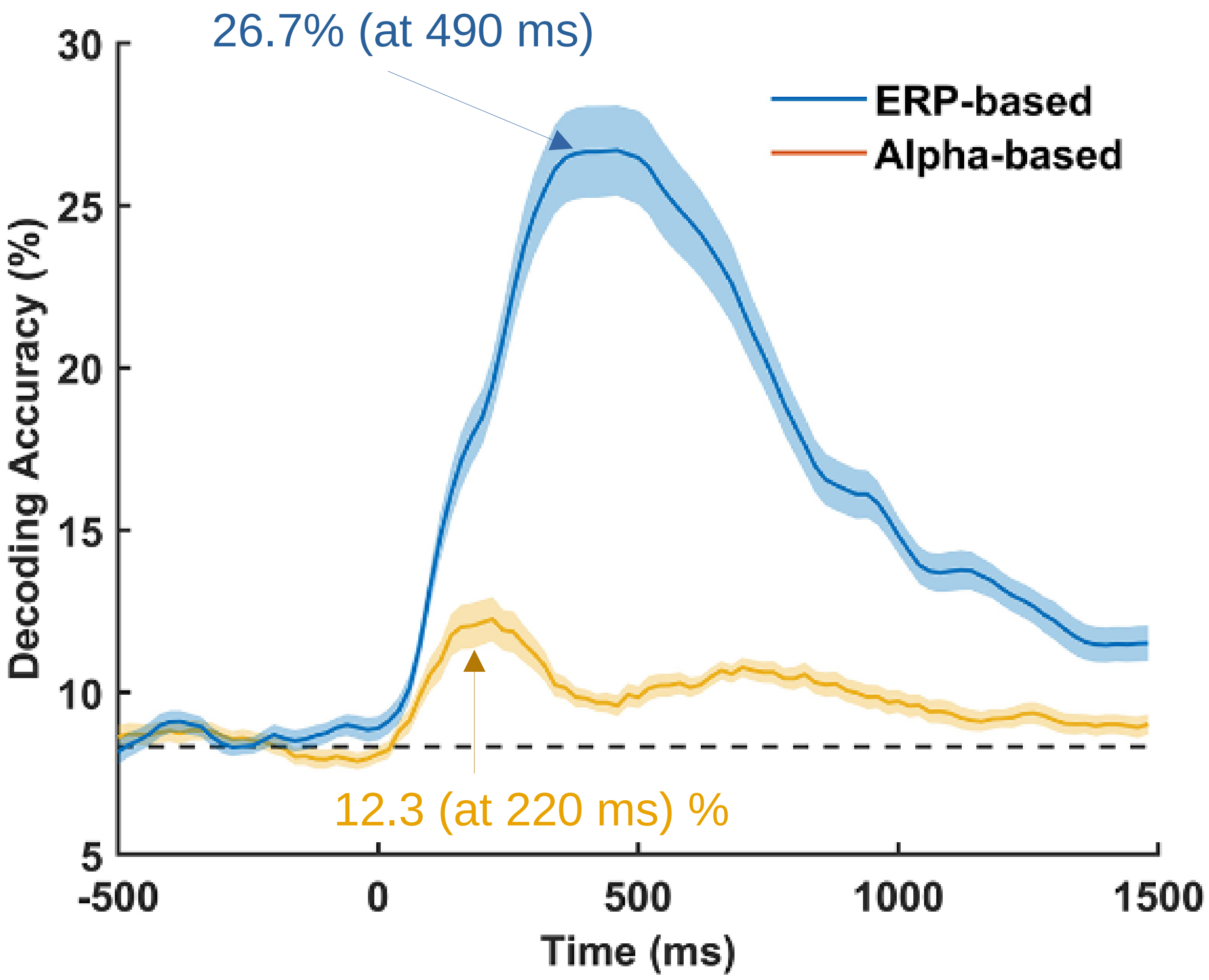
Approach

- Decode SD at each time point to predict numerical magnitudes, for each configuration.
- Compare temporal aspects of decoding accuracy to compare finger configurations
- Compare results with traditional analysis.

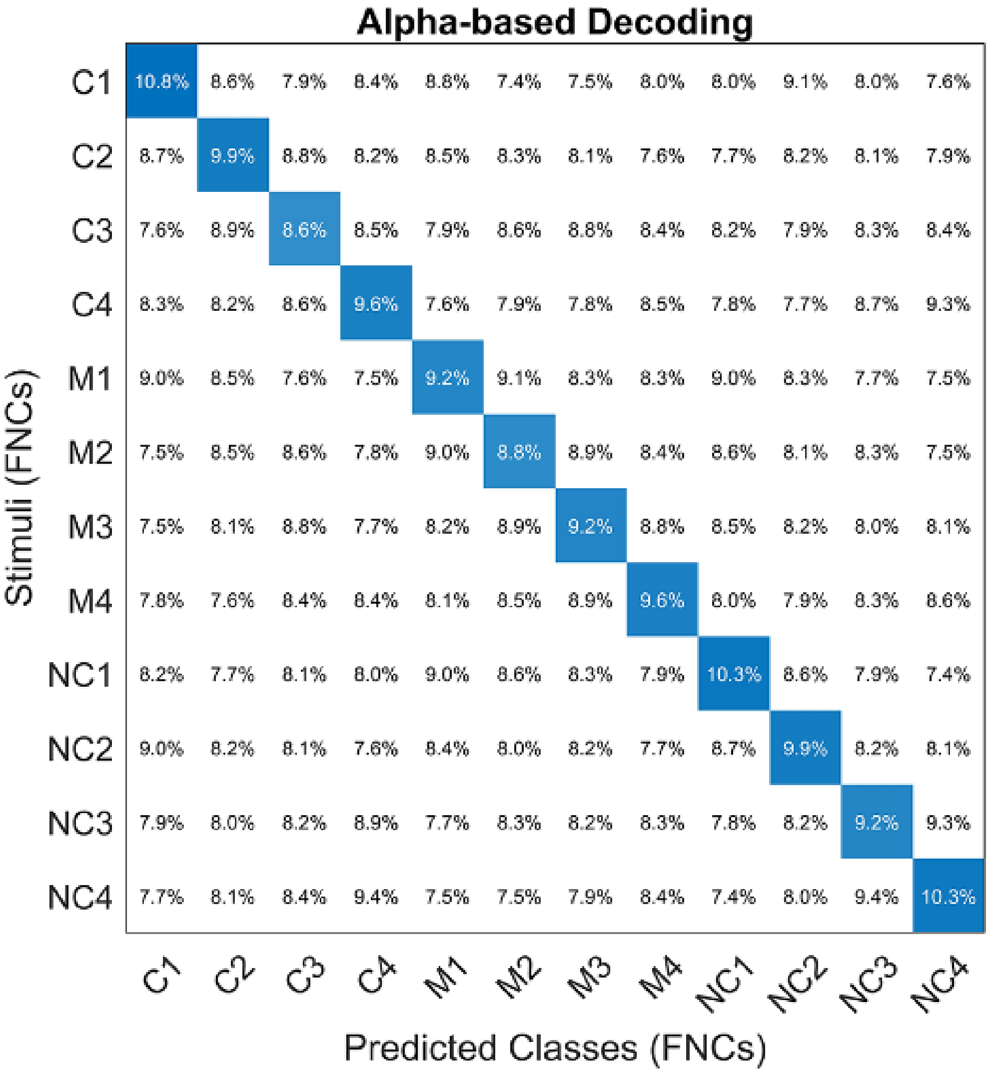


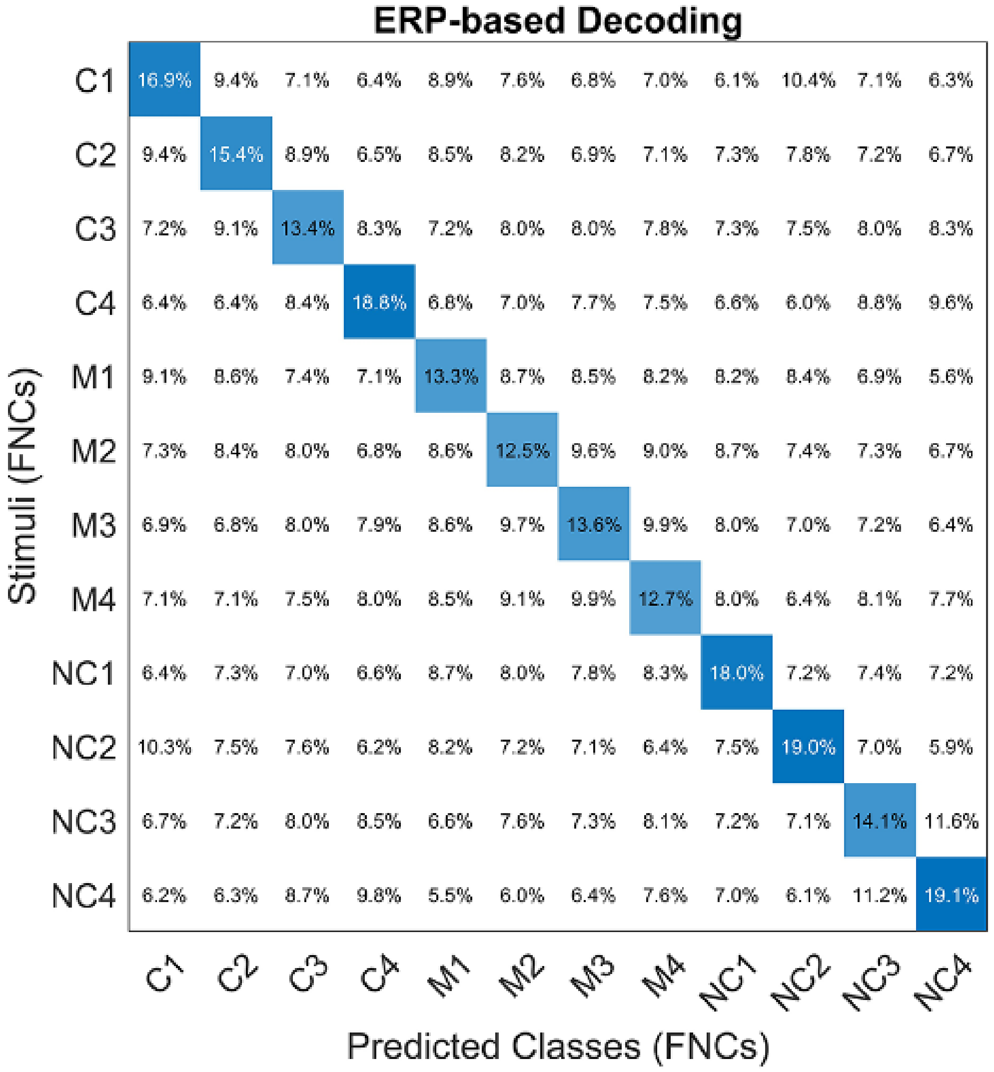
Bae & Luck (2018) Dissociable decoding of spatial attention and working memory from EEG oscillations and sustained potentials, *Journal of Neuroscience*, 38(2), pp. 409-422.

Mean Accuracy of Alpha-Based Decoding and ERP-Based Decoding Averaged Across All 12 FNCs

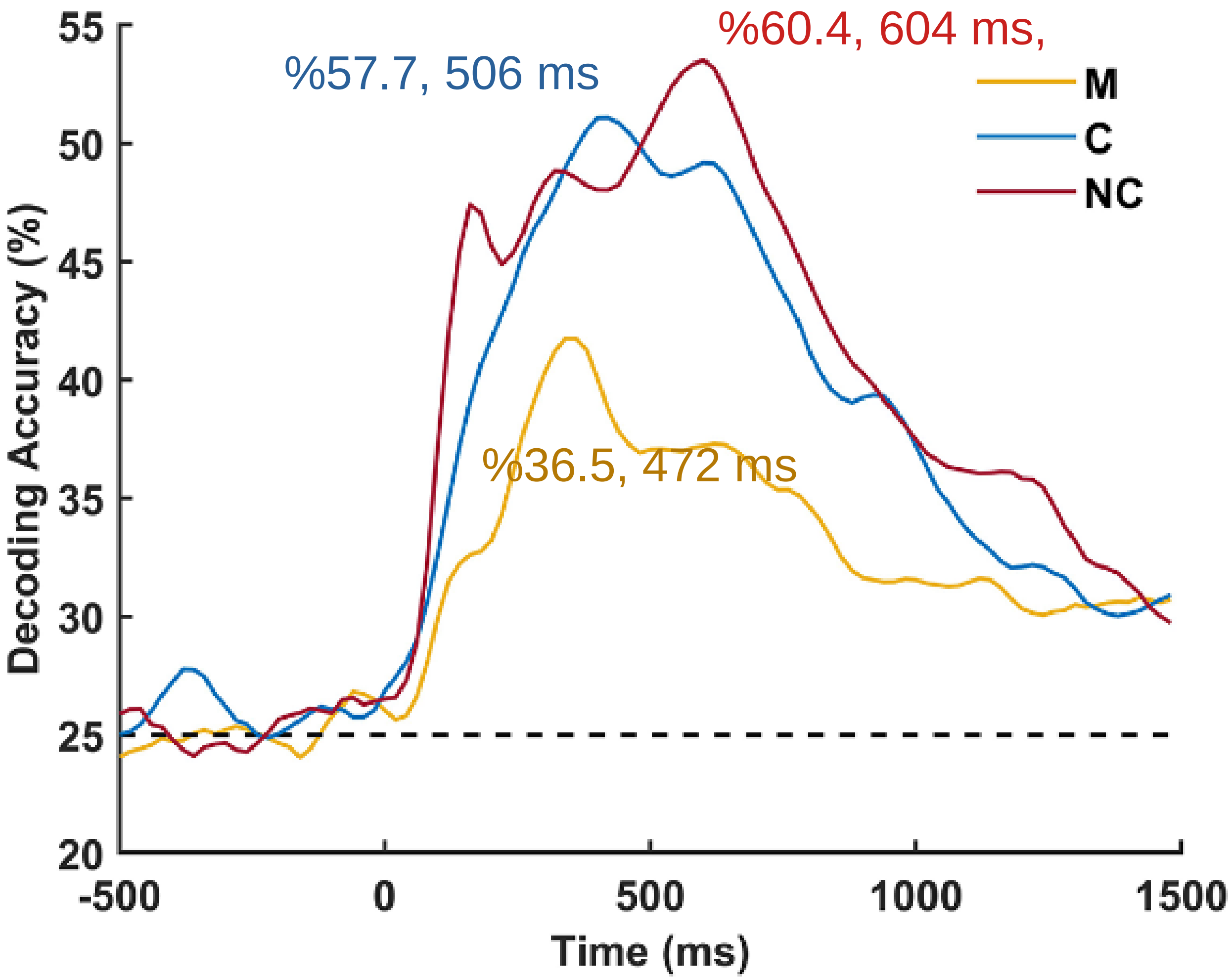


Note. The black horizontal line indicates the chance-level performance (1/12 ≈ 0.083≈ 8.3%). The shaded areas indicate ± 1 SEM.


Alpha-based Decoding
Stimuli (FNCs)
Predicted Classes (FNCs)
C1 C2 C3 C4 M1 M2 M3 M4 NC1 NC2 NC3 NC4
C1 10.8% 8.6% 7.9% 8.4% 8.8% 7.4% 7.5% 8.0% 8.0% 9.1% 8.0% 7.6%
C2 8.7% 9.9% 8.8% 8.2% 8.5% 8.3% 8.1% 7.6% 7.7% 8.2% 8.1% 7.9%
C3 7.6% 8.9% 8.6% 8.5% 7.9% 8.6% 8.8% 8.4% 8.2% 7.9% 8.3% 8.4%
C4 8.3% 8.2% 8.6% 9.6% 7.6% 7.9% 7.8% 8.5% 7.8% 7.7% 8.7% 9.3%
M1 9.0% 8.5% 7.6% 7.5% 9.2% 9.1% 8.3% 8.3% 9.0% 8.3% 7.7% 7.5%
M2 7.5% 8.5% 8.6% 7.8% 9.0% 8.8% 8.9% 8.4% 8.6% 8.1% 8.3% 7.5%
M3 7.5% 8.1% 8.8% 7.7% 8.2% 8.9% 9.2% 8.8% 8.5% 8.2% 8.0% 8.1%
M4 7.8% 7.6% 8.4% 8.4% 8.1% 8.5% 8.9% 9.6% 8.0% 7.9% 8.3% 8.6%
NC1 8.2% 7.7% 8.1% 8.0% 9.0% 8.6% 8.3% 7.9% 10.3% 8.6% 7.9% 7.4%
NC2 9.0% 8.2% 8.1% 7.6% 8.4% 8.0% 8.2% 7.7% 8.7% 9.9% 8.2% 8.1%
NC3 7.9% 8.0% 8.2% 8.9% 7.7% 8.3% 8.2% 8.3% 7.8% 8.2% 9.2% 9.3%
NC4 7.7% 8.1% 8.4% 9.4% 7.5% 7.5% 7.9% 8.4% 7.4% 8.0% 9.4% 10.3%


ERP-based Decoding
Stimuli (FNCs)
Predicted Classes (FNCs)
C1 C2 C3 C4 M1 M2 M3 M4 NC1 NC2 NC3 NC4
C1 16.9% 9.4% 7.1% 6.4% 8.9% 7.6% 6.8% 7.0% 6.1% 10.4% 7.1% 6.3%
C2 9.4% 15.4% 8.9% 6.5% 8.5% 8.2% 6.9% 7.1% 7.3% 7.8% 7.2% 6.7%
C3 7.2% 9.1% 13.4% 8.3% 7.2% 8.0% 8.0% 7.8% 7.3% 7.5% 8.0% 8.3%
C4 6.4% 6.4% 8.4% 18.8% 6.8% 7.0% 7.7% 7.5% 6.6% 6.0% 8.8% 9.6%
M1 9.1% 8.6% 7.4% 7.1% 13.3% 8.7% 8.5% 8.2% 8.2% 8.4% 6.9% 5.6%
M2 7.3% 8.4% 8.0% 6.8% 8.6% 12.5% 9.6% 9.0% 8.7% 7.4% 7.3% 6.7%
M3 6.9% 6.8% 8.0% 7.9% 8.6% 9.7% 13.6% 9.9% 8.0% 7.0% 7.2% 6.4%
M4 7.1% 7.1% 7.5% 8.0% 8.5% 9.1% 9.9% 12.7% 8.0% 6.4% 8.1% 7.7%
NC1 6.4% 7.3% 7.0% 6.6% 8.7% 8.0% 7.8% 8.3% 18.0% 7.2% 7.4% 7.2%
NC2 10.3% 7.5% 7.6% 6.2% 8.2% 7.2% 7.1% 6.4% 7.5% 19.0% 7.0% 5.9%
NC3 6.7% 7.2% 8.0% 8.5% 6.6% 7.6% 7.3% 8.1% 7.2% 7.1% 14.1% 11.6%
NC4 6.2% 6.3% 8.7% 9.8% 5.5% 6.0% 6.4% 7.6% 7.0% 6.1% 11.2% 19.1%

Mean Accuracy of ERP-Based Decoding for Montring (M), Counting (C), and Noncanonical (NC) Configurations

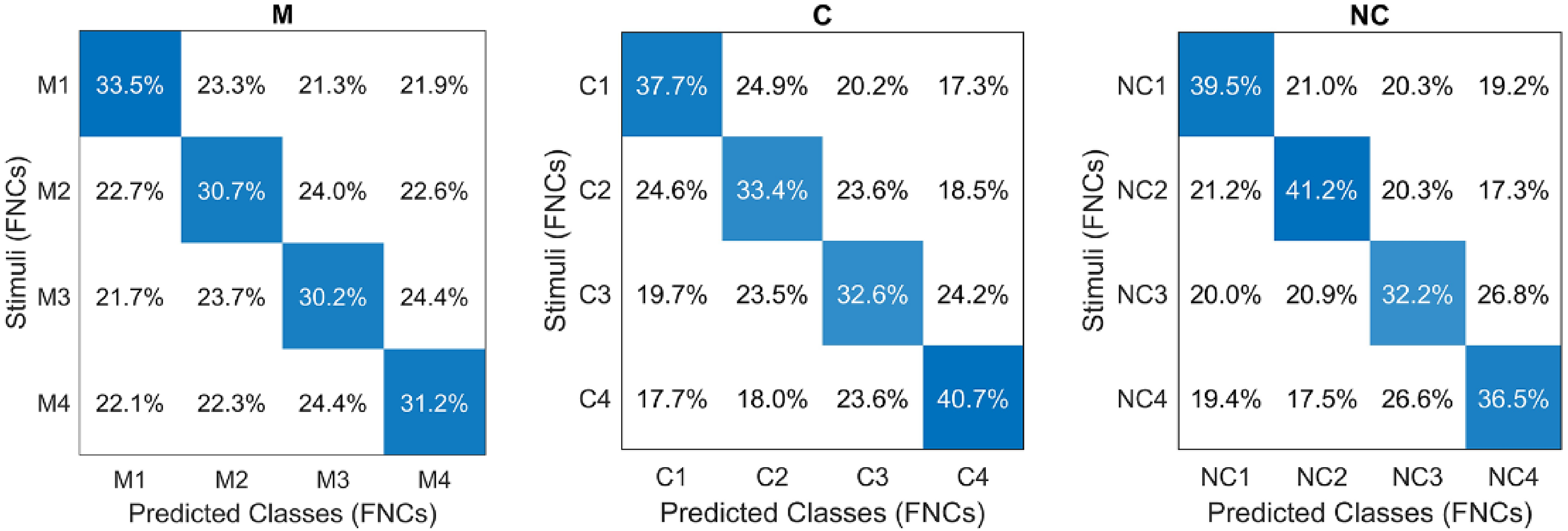


Note. The black horizontal line indicates the chance-level performance (0.25 = 1/4).

Comparison of configuration types:

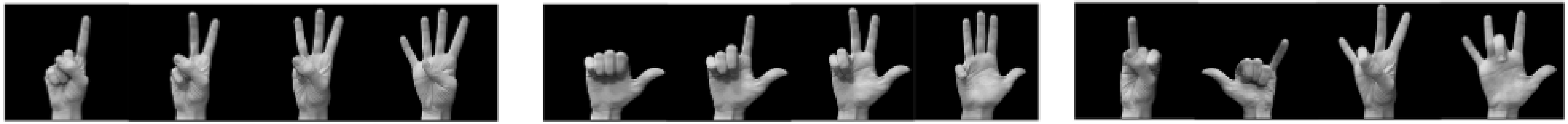
- Rerun the procedure for each conf.
- Compare the temporal aspects of avg. accuracy

The Percentage of Accurate Classifications for Each of the 4 FNCs



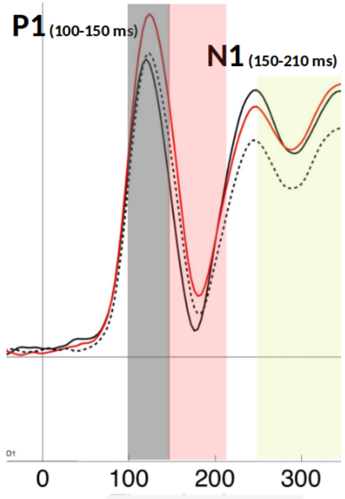
FNC	Average accuracy (%) *M* (*SD*)	Peak accuracy (%) *M* (*SD*)	Peak accuracy time (ms) *M* (*SD*)
Montring (M)	36.5 (6.6)	47.2 (8.9)	445.789 (186.629)
Counting (C)	44.3 (9.6)	57.7 (11.4)	506.316 (184.192)
Noncanonical (NC)	46.3 (10.9)	60.4 (12.6)	515.263 (194.377)

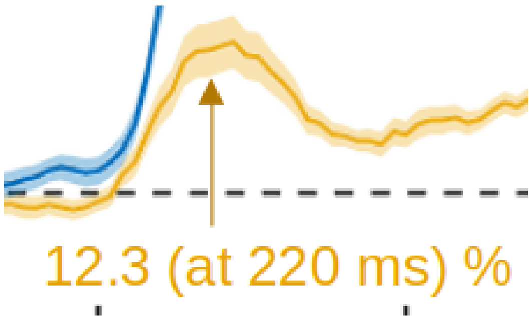
Montring (M) **Counting (C)** **Non-canonical (NC)**



Results & Comparisons

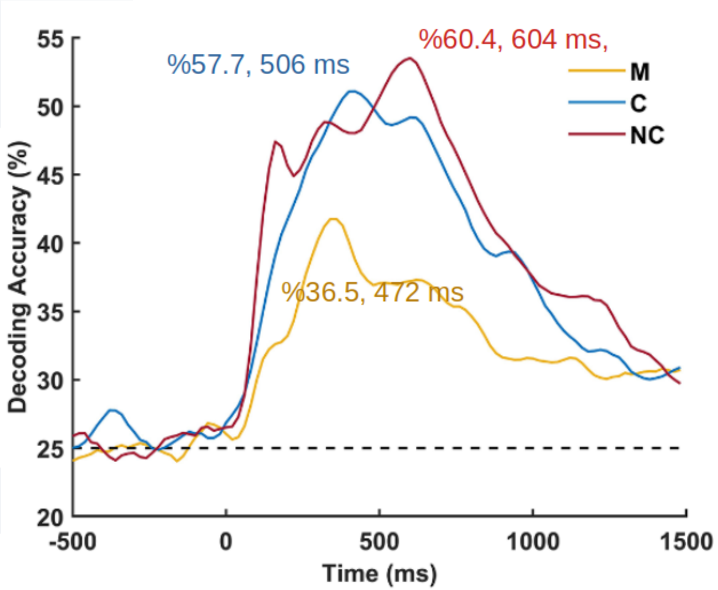
- ERP-based more accurate than Alpha-based
- Early peak for alpha-decoding *might be* associated with the previous P1/N1 effect *(caution due to low acc. level)*

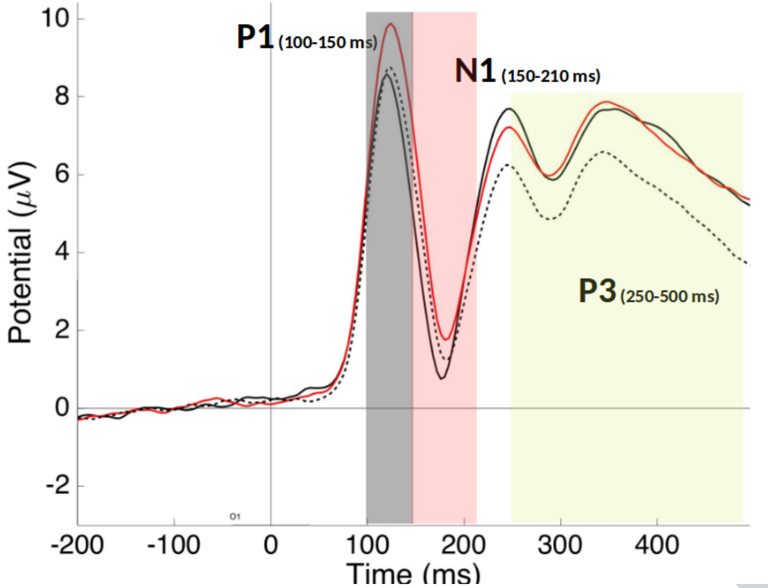




Results & Comparisons

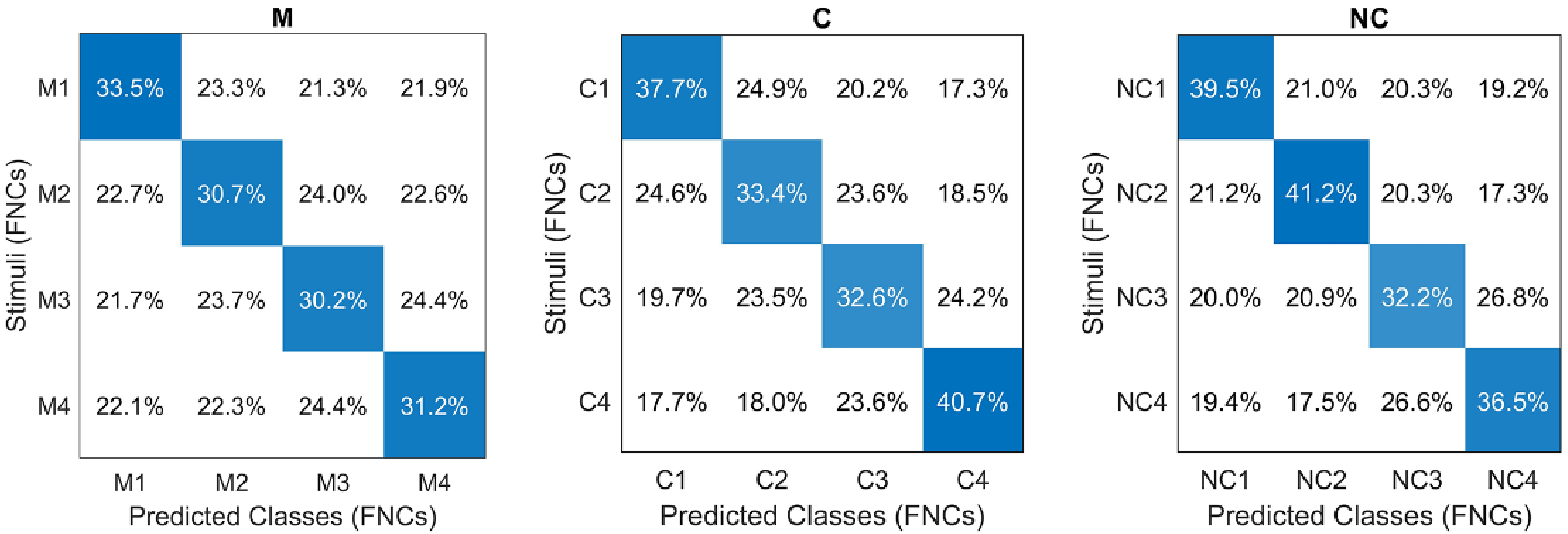
- Decoding better informs when the three conditions diverge in processing (post-500ms).





Results & Comparisons

- Decoding allows a more detailed inspection of differences across conditions.

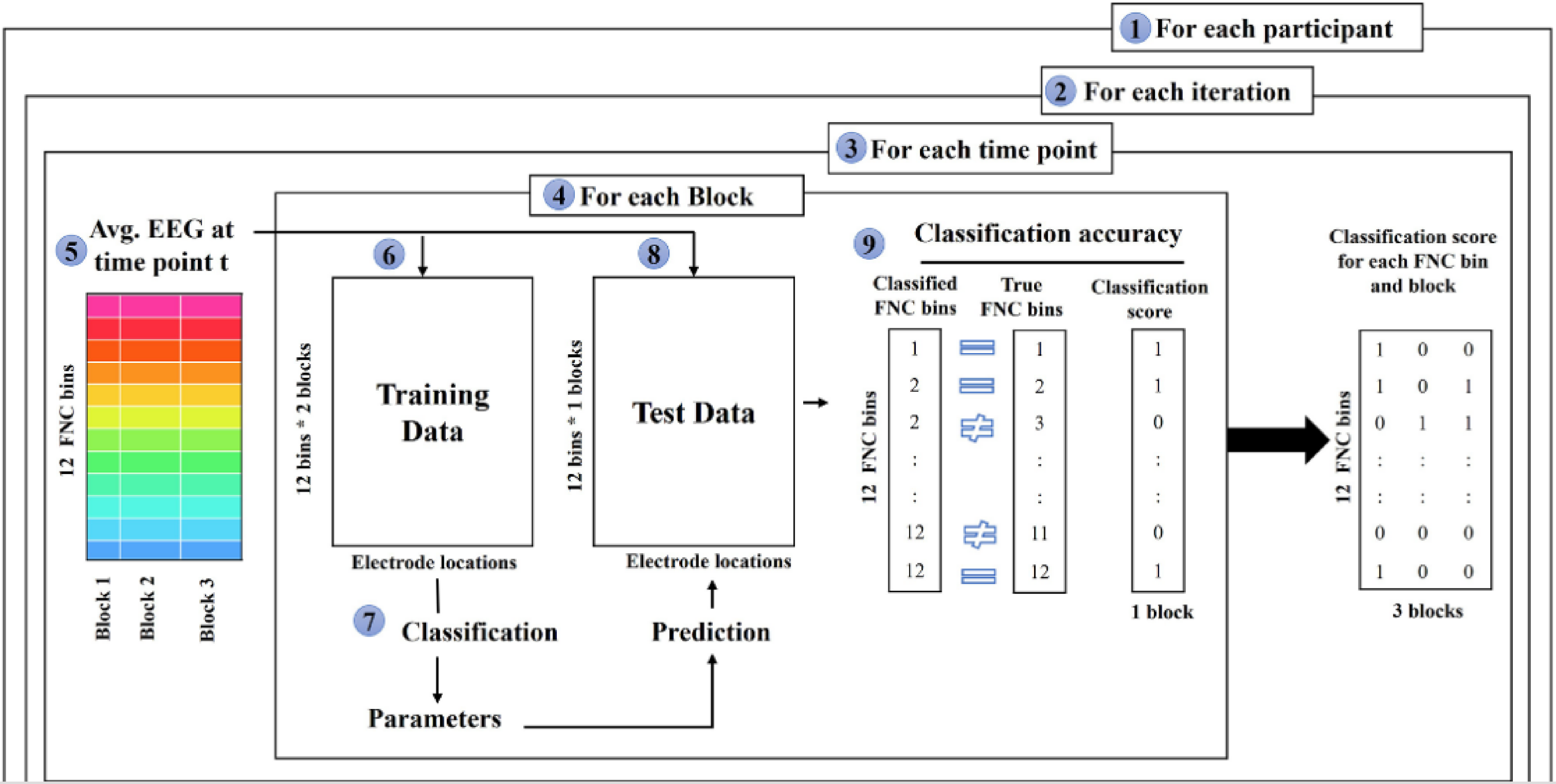


What about other ML algorithms?

Decoding Performance of 6 ML Algorithms Compared

Classifier	MATLAB function	Detail
SVM	*fitcecoc()*	One. vs all, linear kernel
LDA	*fitcecoc()*	One. vs all, linear discriminant, Kernel: normal
NB	*fitcecoc()*	One. vs all
KNN	*fitcecoc()*	One. vs all, k=12 for decoding all FNCs and K=4 for category-specific decoding
DT	*fitcecoc()*	One. vs all
NN	*fitcnet()*	Input: [1,32], FC1: [10,32], FC2: [4,10] or [12, 10], output: [1,1]

Salehzadeh, R., Soylu, F., & Jalili, N. (2023) A Comparative Study of Machine Learning Methods for Classifying ERP Scalp Distribution. *Biomedical Physics & Engineering Express*. https://doi.org/10.1088/2057-1976/acdbd0



Bae & Luck (2018) Dissociable decoding of spatial attention and working memory from EEG oscillations and sustained potentials, *Journal of Neuroscience*, 38(2), pp. 409-422.

4 Comparison Metrics

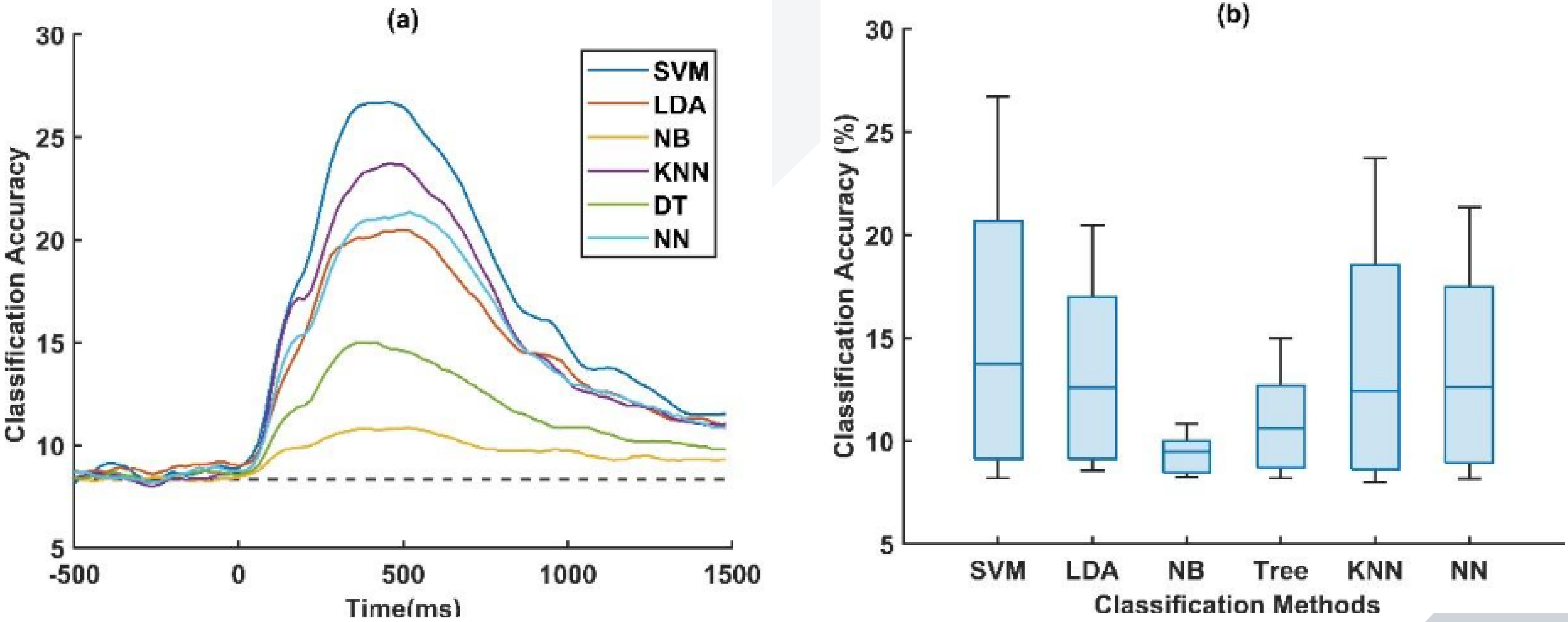
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

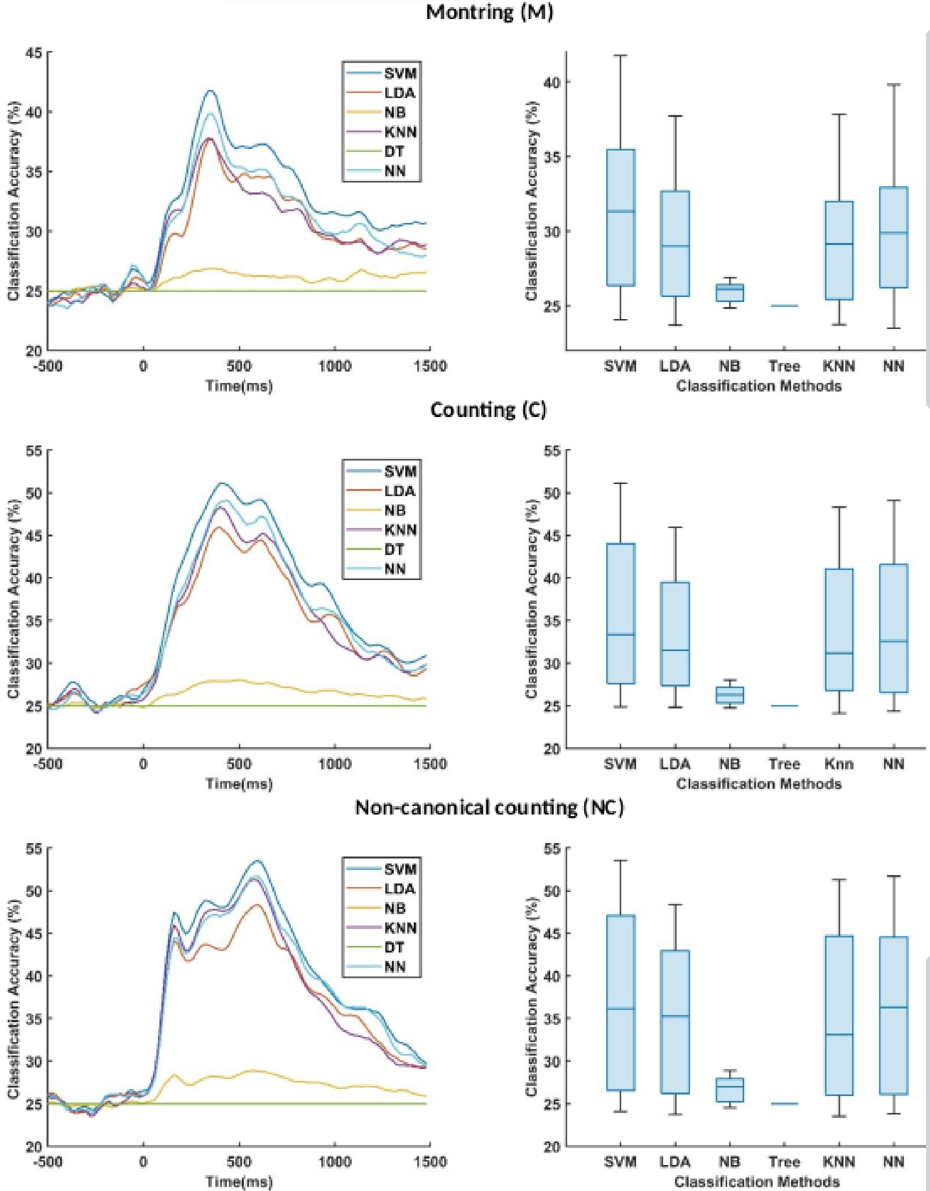
$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

TP: true positives, TN: true negatives
FP: false positives, FN: false negatives


(a)
SVM
LDA
NB
KNN
DT
NN
Classification Accuracy
Time(ms)
(b)
Classification Accuracy (%)
SVM
LDA
NB
Tree
KNN
NN
Classification Methods

Classifier	Avg. Accuracy	Precision	Recall	F-Score
SVM	15.46	15.57	15.46	15.51
LDA	13.45	13.53	13.45	13.49
NB	9.46	10.96	9.46	10.15
KNN	14.13	14.22	14.13	14.18
DT	10.98	11.13	10.98	11.06
NN	13.59	13.62	13.59	13.60

Category specific decoding

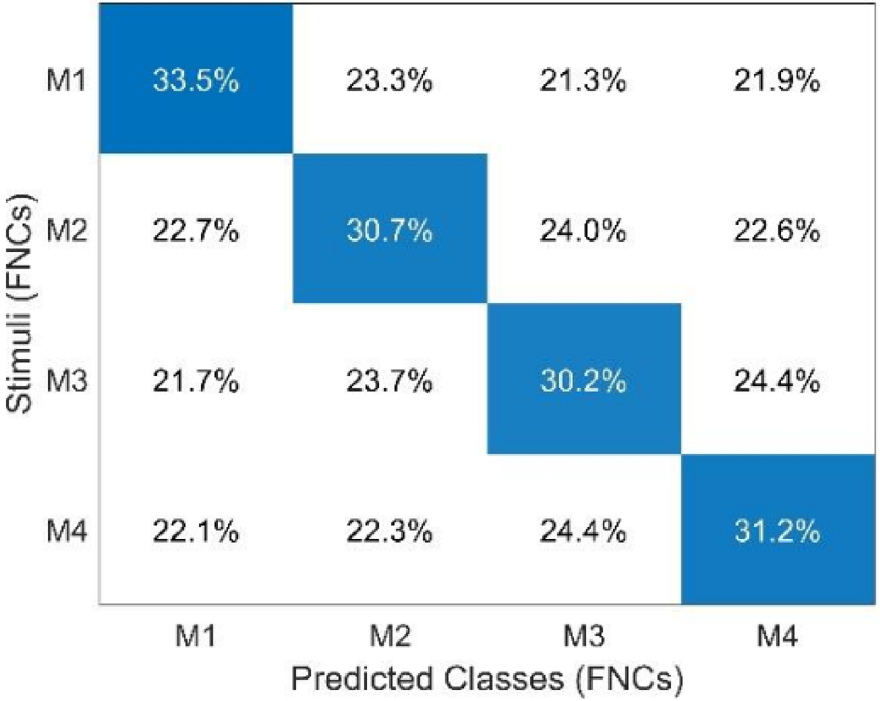


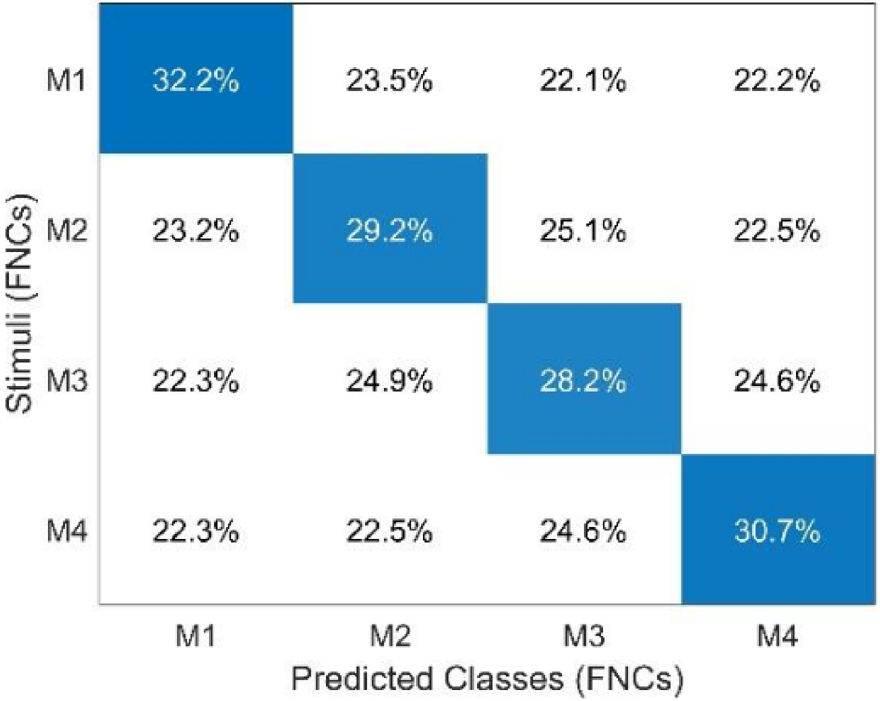
Confusion Matrices

(SVM)

(NN)

Montring (M)

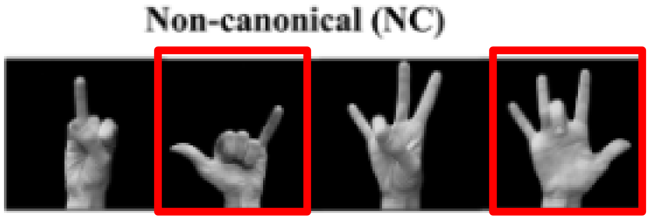




Results

- SVM scores highest across all measures -> most reliable. NN & KNN perform well as well.
- DT close to chance-level in category specific, but better in all-categories → DT better suited to decoding with high num. of classes.
- NC & C had higher accuracy than M across all

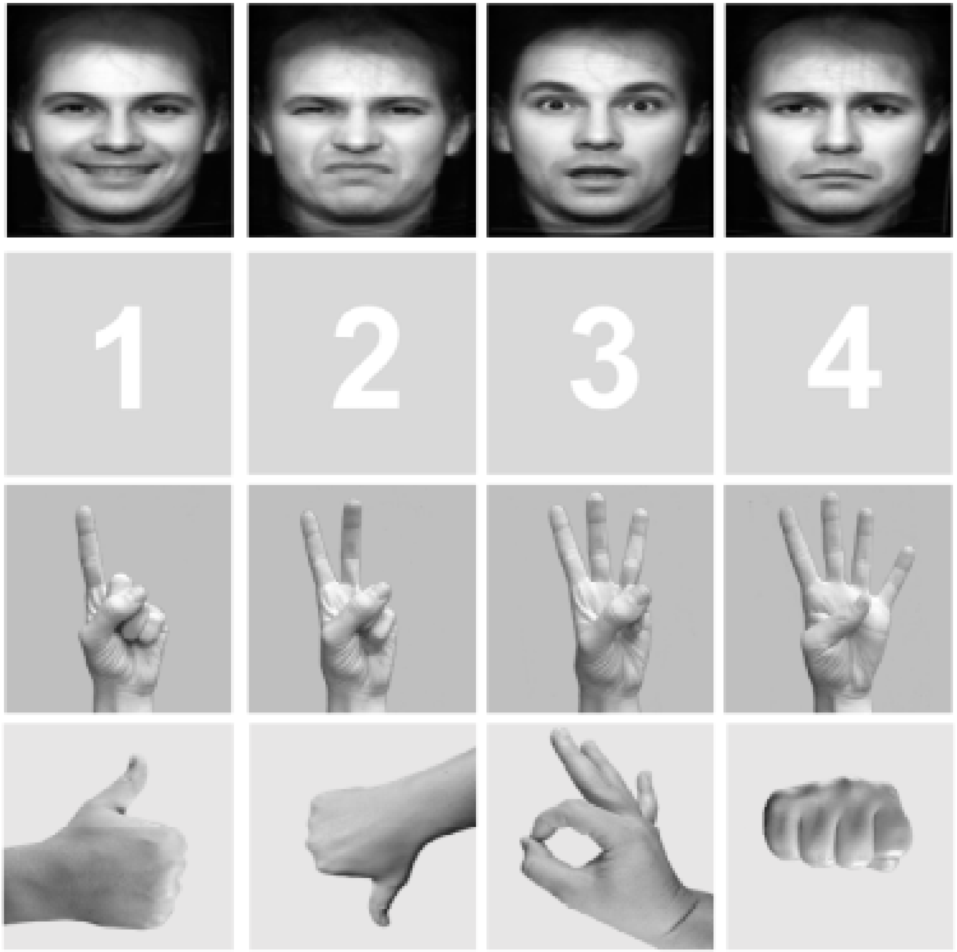
  → automaticity leads to more similar neural patterns
- Highest true positive rates for Non-Canonical, numbers 2 and 4 → most distinguishable



Future Work

- Explore adjustments in preprocessing
- Test different parameter values with each algorithm.
- Test different forms of perceptual, cognitive, and socio-emotional processing --->

Current Study

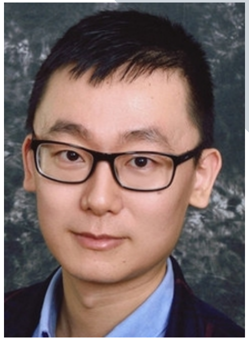


Thank you!

Dr. Roya Salehzadeh
Lawrence Technological Uni.

Dr. Brian Rivera
University of Nebraska-Lincoln

Dr. Nader Jalili
University of Alabama



Dr. Kaiwen Man
University of Alabama

Firat Soylu
The University of Alabama
fsoylu@ua.edu
http://fsoylu.people.ua.edu

Sabancı University
CS Seminar

June 16, 2023