EEG user behavior analysis [May 2]

This text continues the results of Apr 26th. The conclusion was:
We want to present an interpretable model that specifically illustrates the classification results.

User behavior analysis

We test the hypothesis: that classification quality correlates with approximation quality. The approximation quality is the ranked (high accuracy goes first) users according to the best models (Gaussian process classification, Naive Bayes, Logistic regression).

Table 1. Correlation between the order of users after modeling and their behavior

Users	1034	1037	1045	1158	1363	1368	1385	2038	6639	7974	7977	7980	1327	Kendall	**Spearman**
Order of users	**13**	**4**	**2**	**1**	**6**	**11**	**12**	**9**	**5**	**8**	**7**	**3**	**10**	1	**1**
Average time to response (samples)	249	277	206	226	208	197	257	254	287	217	216	251	236	0.08	**0.03**
Standard deviation of time of response	65	55	54	51	49	45	47	54	53	48	64	61	60	-0.05	**-0.06**
Number of changes in responses	8	14	14	12	32	25	12	25	14	41	35	12	9	-0.01	**-0.11**
Average time between first second responses	55	41	69	30	35	27	38	45	37	45	49	38	34	0	**0.05**
Standard deviation of time	93	9	26	11	21	6	12	11	8	9	12	6	15	0.13	**0.19**
Number of incorrect responses	33	6	8	20	10	10	6	14	7	5	24	18	7	-0.04	**-0.04**
Average time to incorrect response	254	251	229	230	209	236	308	277	278	254	262	261	256	0.23	**0.36**
Standard deviation of time	254	251	229	230	209	236	308	277	278	254	262	261	256	0.23	**0.36**
Number of second incorrect responses	91	57	80	97	78	72	57	65	44	55	76	93	42	-0.3	**-0.37**

Set priorities on classification models and select users

Since the number of sample objects is insufficient and the number of features is redundant (about 80 events per user against 452x128 per user's ERP), we continue to find ways to augment the sample set and select features. Two ways of sample set augmentation: combine users and find data sets similar to the given one.

Table 2 shows the list of ordered users. Six users were selected for grouping: 1158, 1045, 7980, 1037, 6639, and 1363. We join users with the highest accuracy to avoid instant overtraining. We may need additional alignment to make a group (signal amplitude and peak dynamic time warping). After the user joins, the model trains the *whole group with reduced accuracy and is tested on each user separately.*

Table 2 collects the results from the computational experiment on Apr 26th. Three basic models were selected according to the obtained accuracy, The Logistic regression model as the simplest one, the Naive Bayesian classification model as the generalization of the LR, and the Gaussian process classification model as the generalization of these two. The deterministic part of these models is the linear combination of features. Further, we will investigate classes of Space-State models under the plausibility of Gaussian processes models.

Table 2. Prioritized models and ordered users. In the recognition task with Old versus New words, the events with incorrect answers were dropped out.

User	**1158**	**1045**	**7980**	**1037**	**6639**	**1363**	7977	7974	2038	1327	1368	1385	1034	**mean**
GP rbf	0.99	0.92	0.85	0.92	0.9	0.97	0.78	0.92	0.59	0.98	0.8	0.45	0.64	**0.8**
GP quad	0.99	0.92	0.85	0.92	0.9	0.97	0.78	0.92	0.59	0.98	0.8	0.45	0.64	**0.8**
Logistic Regr	**0.97**	**0.99**	**0.98**	**0.9**	**0.9**	**0.77**	**0.84**	**0.74**	**0.89**	**0.59**	**0.5**	**0.7**	**0.51**	**0.79**
Linear SVM	0.97	0.99	0.98	0.92	0.92	0.8	0.85	0.78	0.9	0.61	0.41	0.59	0.56	**0.79**
Naive Bayes	**0.97**	**0.99**	**0.98**	**0.92**	**0.92**	**0.81**	**0.85**	**0.78**	**0.9**	**0.57**	**0.49**	**0.67**	**0.53**	**0.79**
GP matern	0.99	0.92	0.85	0.92	0.9	0.97	0.78	0.92	0.59	0.98	0.8	0.44	0.54	**0.79**
Gaussian Pr	**0.97**	**0.99**	**0.98**	**0.9**	**0.84**	**0.78**	**0.83**	**0.74**	**0.89**	**0.57**	**0.52**	**0.69**	**0.5**	**0.78**
Neural Net	0.97	0.99	0.98	0.9	0.88	0.79	0.84	0.74	0.89	0.54	0.52	0.69	0.51	**0.78**
GP linear	0.99	0.9	0.84	0.9	0.89	0.97	0.74	0.9	0.5	0.98	0.78	0.49	0.65	**0.78**
Nearest Nbr	0.97	0.99	0.98	0.9	0.78	0.75	0.79	0.72	0.87	0.57	0.55	0.74	0.5	**0.77**
SVM	0.99	0.9	0.83	0.9	0.83	0.97	0.74	0.82	0.53	0.98	0.78	0.46	0.67	**0.77**
AdaBoost	0.94	0.93	0.98	0.82	0.82	0.67	0.76	0.68	0.78	0.52	0.56	0.6	0.56	**0.74**
Decision Tree	0.96	0.97	0.97	0.82	0.88	0.6	0.76	0.64	0.72	0.54	0.5	0.55	0.47	**0.72**
Random Forst	0.93	0.9	0.97	0.82	0.82	0.7	0.85	0.64	0.87	0.47	0.45	0.45	0.54	**0.72**
Quadratic DA	0.64	0.75	0.78	0.59	0.74	0.46	0.58	0.54	0.62	0.47	0.48	0.59	0.56	**0.6**
RBF SVM	0.54	0.46	0.54	0.5	0.56	0.47	0.59	0.52	0.35	0.61	0.38	0.59	0.6	**0.51**
mean	**0.92**	**0.91**	**0.9**	**0.85**	**0.84**	**0.78**	**0.77**	**0.75**	**0.72**	**0.69**	**0.58**	**0.57**	**0.56**	

Warning. The results are *unstable*. The digits now are not trusted (even the accuracy of logistic regression) due to insufficient sample size and a redundant number of features.
Unstable means it easily drops to 0.5 when the data change a little.

The electrode selection

We have two methods of electrode selection: expert and machine learning selection. The second we discuss is in the Feature selection section below. For the first one, the recommended list of electrodes is D7 (F7), C21 (Fz), C7 (F8), A1 (Cz), D31 (P7), A19 (Pz), B11 (P8), A15 (O1) is an approximation from the 10-20 to the BioSemi electrode placement system.

Figure 1 compares the time series for eight expert-recommended electrodes for one user. The hypothesis is that each stimulus evokes a similar ERP. The user's data (events, electrodes, and time samples) was averaged over events. Then the time series were smoothed independently with a Gaussian window of 15 samples. Figure 2 shows that the deviation is high, while seems to be time-stationary.

[The rest users' plots are in the pdf here](), 13 pages. The users are really different in their ERPs.
Not sure how their mixing before feature selection helps.

Also, for each electrode separately, see the YouTube video here. The ERP signals captured from 8 recommended electrodes, each user separately, only Recognition task, event Old word, the events go consequently. https://www.youtube.com/watch?v=ILVczr7tslo. *Please set the video quality 720p.*

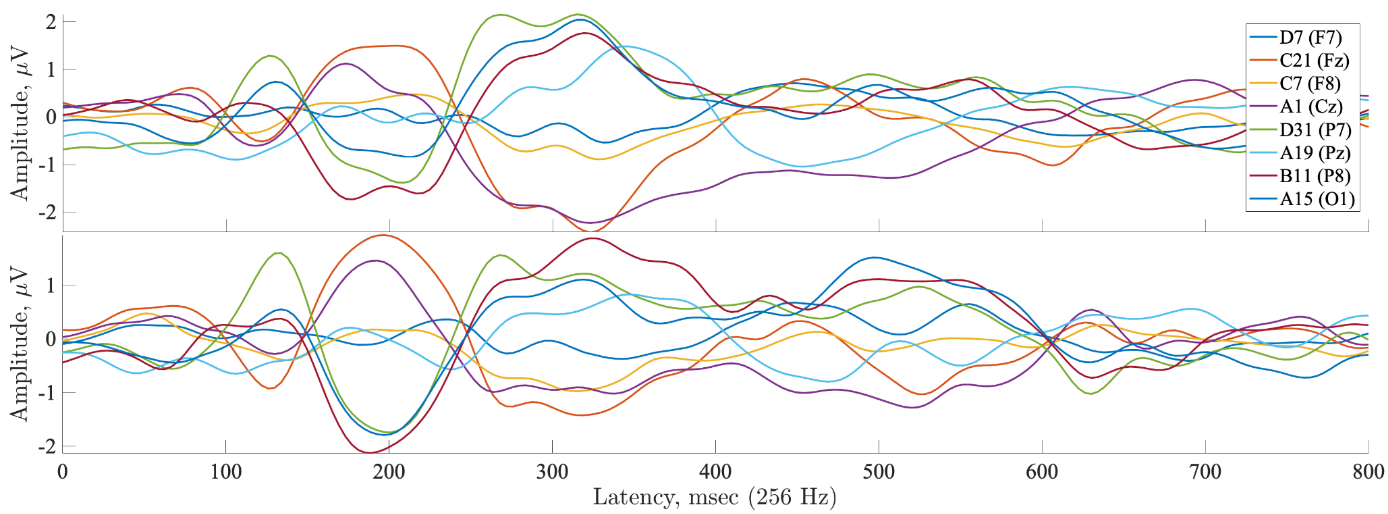

Figure 1. An illustrative example of ERP of over one user for two classes for the recognition task, class Old word at the top and class New word at the bottom, the incorrect answers are dropped.

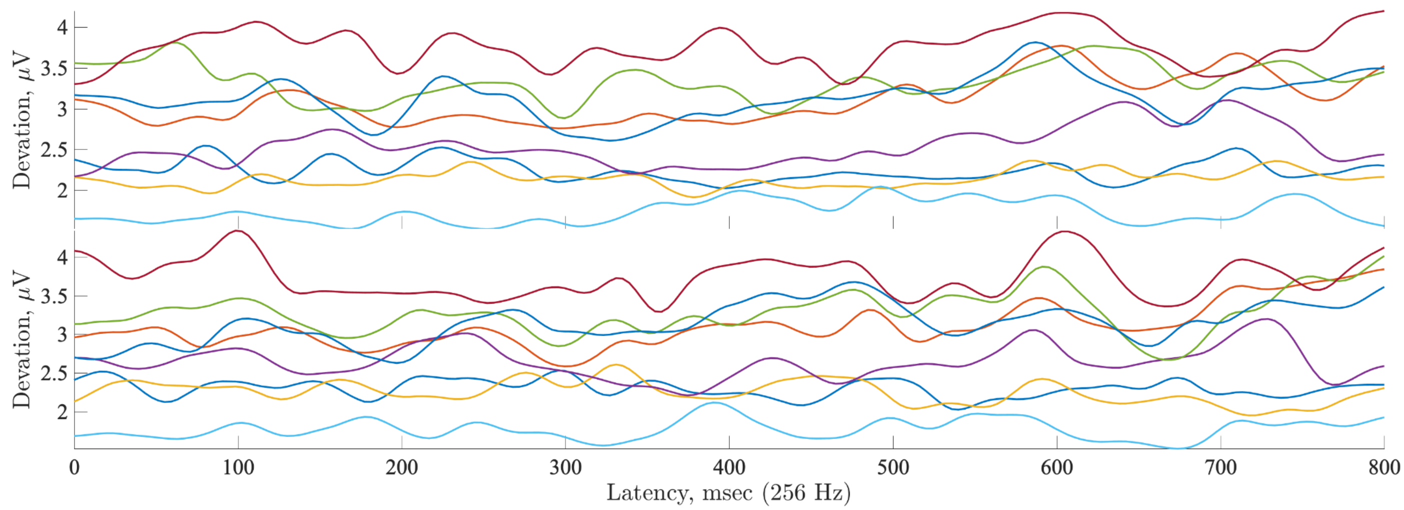

Figure 2. The standard deviation of the corresponding ERP signals from Figure 1.

Overtraining, AUC, and accuracy

The main problem is the small number of samples. Typically each class for a user contains 44 versus 39 samples (training and test included). The visualization above shows that joining users into one dataset is difficult because their ERPs differ. The feature space is 128 x 542. So it is possible to show **any** result that will be highly insatiable given new data.

Below, Figure 3 a), b), and Figure 4 show the first three selected users (according to the best classification quality), the first six selected users, and all 13 users selected consequently.

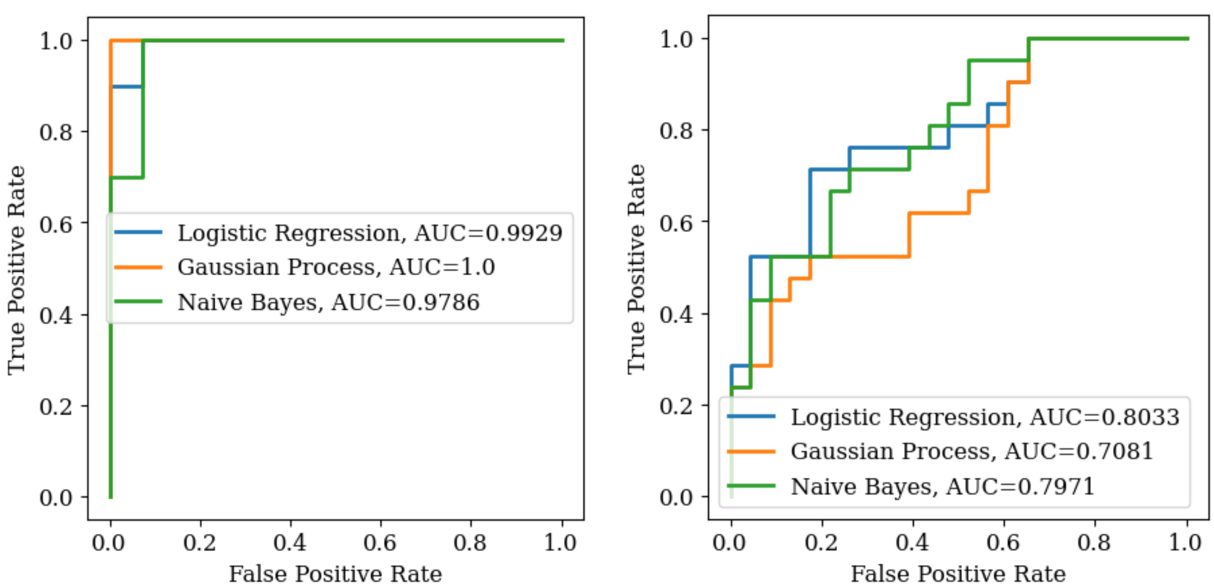

Figure 3. The ROC for the users, selected according to the best accuracy, Task Recognition, Old versus New words, correct responses, left a) three users 1158, 1045, 7980; right b) six users +1037, 6639, 1363.

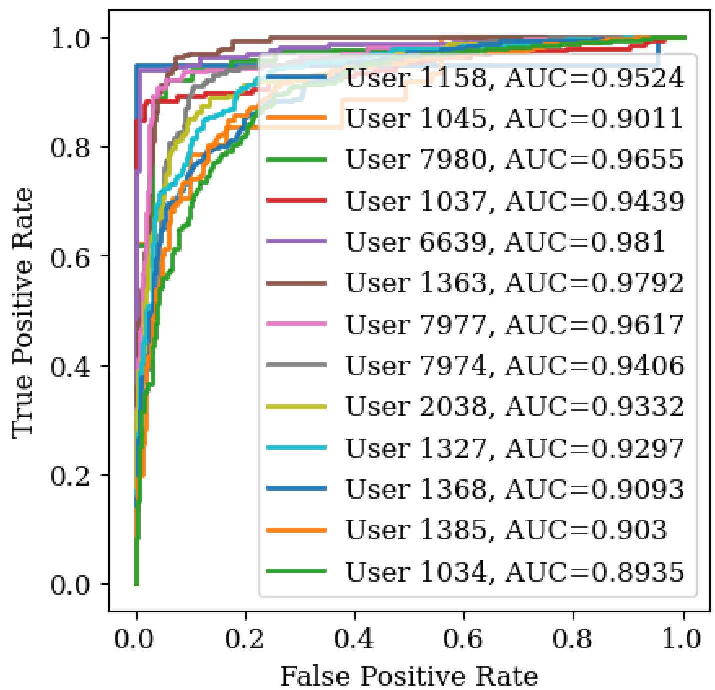

Figure 4. The ROC for the accumulated users. The order is 1158, 1045, 7980, 1037, 6639, 1363, 7977, 7974, 2038, 1327, 1368, 1385, and 1034. Only the Logistic regression model was trained. It receives consequently the first, the first two,... etc., the first all.

Datasets to select from

Since the linear model is accepted as basic, the number of features is the only structure parameter. It depends on the sample size, which is insufficient now. So we run this feature selection procedure on various data sets, which our experts name similar to the given data.
Below is the list of open EEG ERP collections. Some of these datasets might be used to compare the structure of selected features. We need a dataset with many similar two-class EEG ERP events to analyze the structure of extracted features and, furthermore, to analyze the error of our forecasting model.

1. MNE tutorial datasets
    a. Example datasets
2. Mobile BCI dataset: Scalp- and ear-EEGs with ERP and SSVEP paradigms while standing, walking, and running
    a. ArXiv
    b. Data
    c. paper-with-code
3. Kaggle EEG data (2): From sensory tasks in Schizophrenia 9 GB with ERP
    a. Kaggle
    b. Get started
    c. First version of data

4. The Nencki-Symfonia EEG/ERP dataset: First, data validation confirmed the acceptable quality of the obtained EEG signals. Typical event-related potential (ERP) waveforms were obtained, as expected, for attention and cognitive control tasks (i.e., N200, P300, N450).
    a. FTP
    b. Paper
5. BCI Competition IV 2008: Motor imagery, the hand movement direction in MEG
    a. HTML links to five datasets
6. International BCI Competition 2022 Five tracks: Few-shot EEG learning, Microsleep detection from single-channel EEG, Imagined speech classification, Upper-limb movements decoding in a single-arm, EEG(+Ear-EEG)-based ERP detection during walking
    a. Links
7. NeuroKit2 Datasets: A Python toolbox for neurophysiological signal processing. NeuroKit includes datasets that can be used for testing.
    a. Links
    b. Paper
8. Thinking out loud, an open-access EEG-based BCI dataset
    a. GitHub Nature
9. SEED Dataset: A dataset collection for various purposes using EEG signals
    a. Links
10. DEAP Dataset: A Dataset for Emotion Analysis using EEG, Physiological, and Video
    a. Links
11. OpenNeuro: Search EEG portal
    a. Search
12. Neiry-demons: Detection of ERP presence after stimulus.
    a. Gin

Search through

https://openneuro.org/

https://bids.neuroimaging.io/

Priorities over feature extraction methods

To perform an exhaustive search over these methods, their superpositions, and hyperparameters, list the feature extraction methods.

1. Time-averaging features (recommended)
2. Temporal PCA analysis (or SSA for Singular Spectrum Analysis)
3. Multichannel SSA
4. Spatial-time SSA
5. Graph SSA (Graph Laplacian)

6. ERP features, time, and other MNE-generated features
7. Parametric smoothing, Approximation model parameters as features

8. Spectral features
9. Multitaper, spectral density estimation features
10. Coherogram
11. Cross-spectral matrix

Some methods from papers and tools

1. FFT
2. Band Power
3. Synchronization Likelihood
4. Petrosian and Higuchi Fractal Dimensions
5. Hjorth Parameters
6. Detrended Fluctuation Analysis
7. Sample Entropy
8. Lempel-Ziv Complexity
9. Cross-Correlation Coefficient

Appendix 1

Links to the UMN presentation

Link to the draft of WPI presentation
https://docs.google.com/presentation/d/1Bu434Rrapdv7KAKYGPYvJiihB94KQri4uNalY_dyztk/edit?usp=sharing

Link to the full UMN presentation
https://docs.google.com/presentation/d/1P9MK2TsfpWimP7KMa9xVWMEXr9lHJw9y/edit#slide=id.g21ee50f6d64_2_253

Citation of timing from the Schres paper

Relational Memory Function in Schizophrenia: Electrophysiological Evidence for Early Perceptual and Late Associative Abnormalities // Schizophrenia Research
By Scott R Sponheim et al.
https://doi.org/10.1016/j.schres.2023.02.019
See also https://doi.org/10.1093/arclin/acx082

where amplitudes were largest. For encoding we quantified P1 at PO9/PO10 (110-160ms), P2 at PO7/PO8 (240-320ms), and the late negative potential (LNP) at AF7 (500-800ms). For recognition we quantified P1 at PO9/PO10 (110-160ms), P2 at PO7/PO8 (200-410ms), P3 at PO7/PO8 (410-650ms), and the LNP at AF7 from 500-800 ms during item-specific recognition and 400-750 ms during associative recognition. To further inspect the LNP, individual subject's

Low-frequency band-pass filter is needed?

The last thought of 1158, at Electrode index 103, D7 (F7). The task Recognition. Since the scend or decay goes more than one Event (1sec = 256 samples), the low band pass filter is recommended.

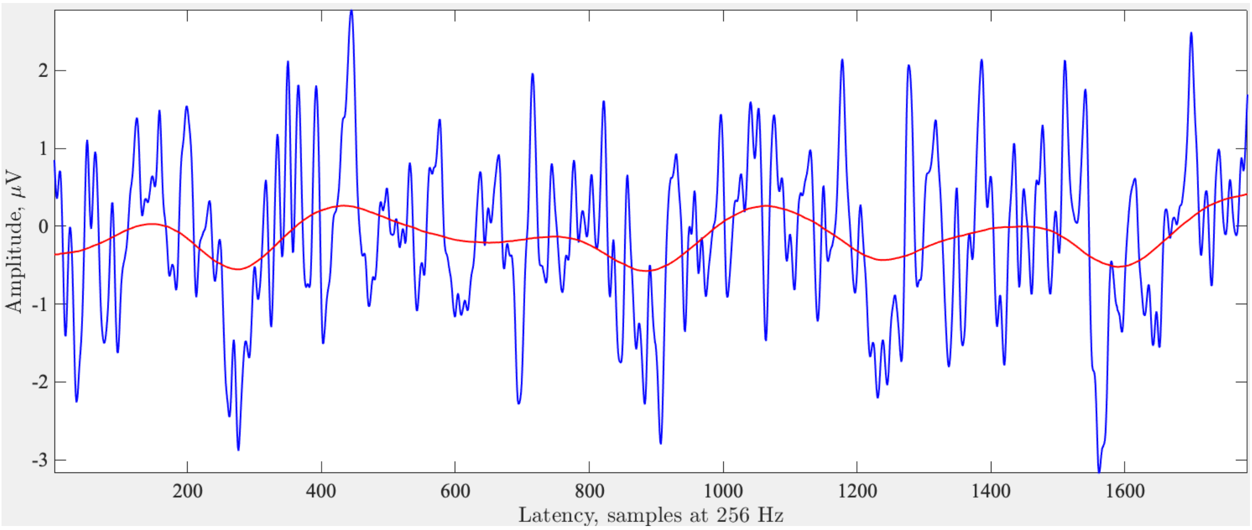

Figure 5. The blue line is Gaussian 15 samples, and the red line is Gaussian 300 samples smoothing.

Appendix 2

Toolboxex, useful for EEG Event-Related Potential feature extraction

Python EEG ERP toolboxes

1. PyRiemann, the main toolbox in this project https://github.com/pyRiemann/pyRiemann
2. https://pyriemann.readthedocs.io/en/latest/auto_examples/index.html#classification-of-erp
3. Event-Related Potentials; it explains how to extract Event-Related Potentials
4. NeuroKit2: The Python Toolbox for Neurophysiological Signal Processing Stress on ECG, slightly on EEG, not ERP, provides a comprehensive suite of processing routines for a variety of bodily signals (e.g., ECG, PPG, EDA, EMG, RSP) https://doi.org/10.3758/s13428-020-01516-y
5. PyEEG https://code.google.com/archive/p/pyeeg, the Reference Guide shows poor list of functions
6. HTNet https://github.com/BruntonUWBio/HTNet_generalized_decoding
7. EEGNet https://github.com/vlawhern/arl-eegmodels/tree/master/examples
8. Gumpy does not downloads data automatically, no way

GP and Space-State Models

9. State Space Models, the latest one S4, https://github.com/srush/annotated-s4/

10. GPPython, for Gaussian process modeling (now the sklearn models are used)
11. GPy, Gaussian Process framework in Python https://gpy.readthedocs.io/en/deploy/

Matlab Toolboxes

1. Signal processing, etc. goes with Matlab
2. EDF file analyzer https://www.mathworks.com/help/signal/ref/edffileanalyzer-app.html
3. EEGLAB
4. FieldTrip https://www.fieldtriptoolbox.org/
5. https://www.mathworks.com/discovery/feature-extraction.html
6. https://en.wikipedia.org/wiki/Coherence_(signal_processing)
7. The BBCI Toolbox is a Brain-Computer Interface (BCI) toolbox https://github.com/bbci/bbci_public
8. ERP_Connectivity_EMG_Analysis https://github.com/EsiSeraj/ERP_Connectivity_EMG_Analysis

Search through the list of papers with code

https://paperswithcode.com/search?q_meta=&q_type=&q=EEG