<u>BCCL - Spring'21 Honours Report</u>

Tanmay Pathak
2018102023

The following is the Project Title and description which was set at the start of the semester.

Project Title: Modelling Simultaneous EEG-fMRI

Project Description: To work on the EEG processing pipeline to extract clean EEG time-series signals and reconcile them with the BOLD fMRI signals recorded simultaneously. One space of correspondence we would explore is the graph spectral embedding space. We would explore these ideas on publicly available datasets.

The outcomes for this semester were

1. Performed literature review on papers describing Simultaneous EEG-fMRI analysis pipelines and compiled relevant datasets.
2. Used Graph Fourier Transform to learn the network harmonics of the brain structural connectivity. Explored the strongly activated network harmonics/brain regions using sliding window analysis.
3. Explored the theory behind EEG Microstates.
4. Familiarised myself with the EEG microstate analysis toolbox

<u>Literature review and database compilation</u>

In the previous semester, we started by looking at methodologies used for modelling simultaneous EEG and fMRI data. Publicly available datasets associated with these papers were also compiled.

<u>For an EEG dataset, the desired qualities are</u>

1) HIGH sampling frequency
2) HIGH sampling resolution
3) Pre-processed dataset
4) Channel map PRESENT
5) No. Of Channels - 32 and 64 are the most common

<u>For a fMRI dataset, the desired qualities are</u>

1) Sequencing method - EPI (Echo Planar Imaging) : gives complete image from a single data sample. It is a technique widely adapted in fMRI imaging to shorten encoding duration and increase temporal resolution
2) HIGH TR (Repetition Time)
3) LOW TE (Echo Time)

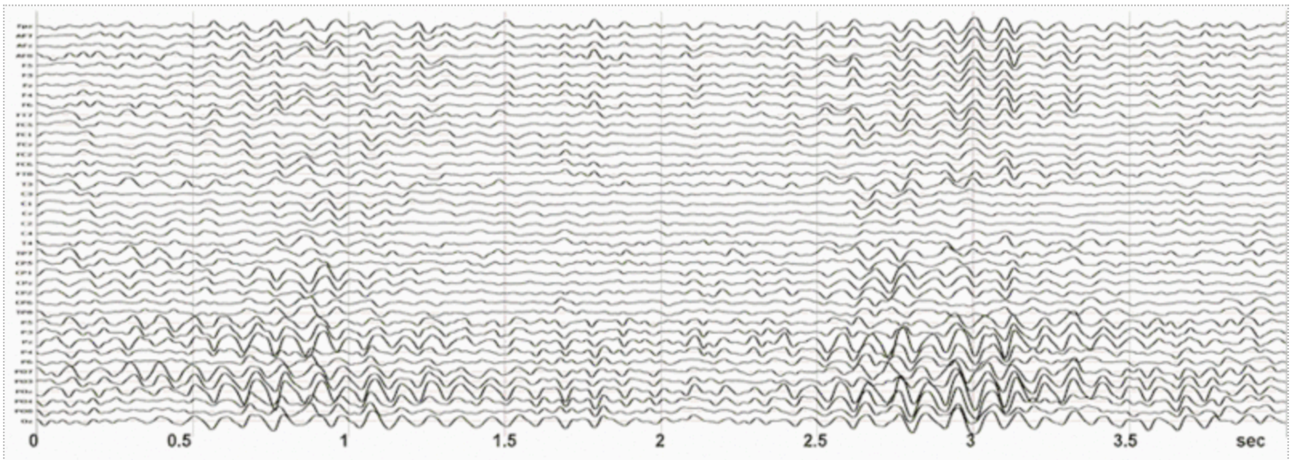
The above mentioned desired characteristics were used to filter datasets and compile the most relevant datasets. <u>LINK to datasets</u>

EEG Microstates - Theory

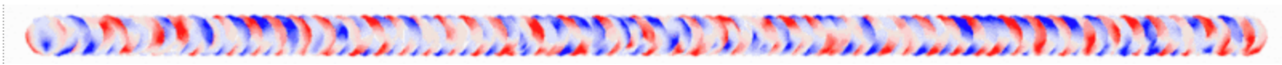
EEG Microstates allow for a **sparse characterisation** of spatial and temporal features of large scale brain network activity.

EEG Microstates, also nicknamed as “the atoms of thought”, are transient, patterned and quasi stable states of EEG.

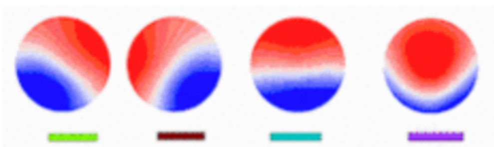
- Transient: They exist only for a short time duration (Durations of microstates during spontaneous task-free resting EEG on average are in the range of 70 to 125 milliseconds)
- Patterned - Observable and noticeable
- Quasi Stable - Global topography is fixed, but strength might vary and polarity invert



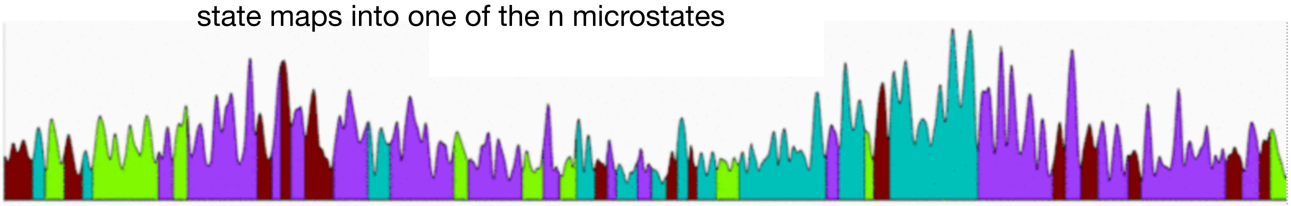
Consider this EEG time series with 32 electrodes sampled at every ‘x’ seconds. For this time series we have ‘x’ EEG states which are [32x1] vectors.



We now have a topographical map corresponding the EEG states.

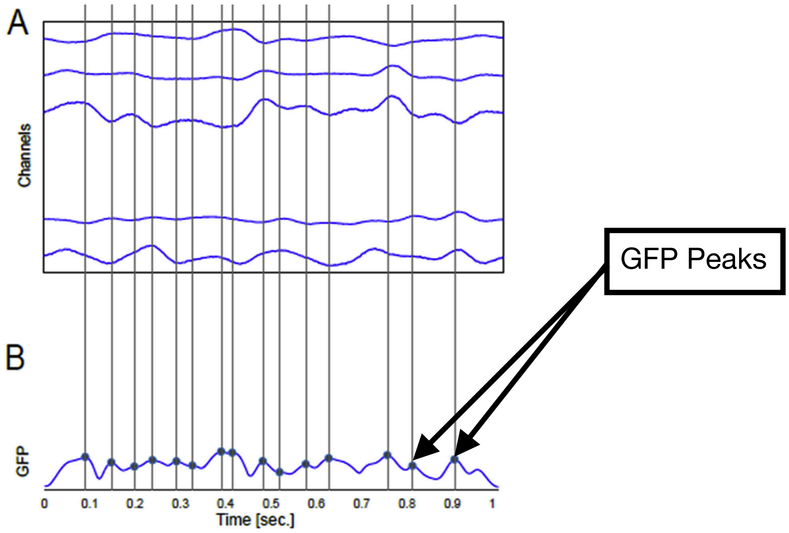


We start with choosing n (here = 4) random template maps. Then we use the k-means clustering algorithm to cluster or classify the EEG state maps into one of the n microstates



After the EEG states have been assigned to a microstate, we see the EEG time series also classified to different (n) microstates. Here we see only one channel data as the above series only shows the GFP

GFP (Global Field Power) - quantifies the amount of activity at each time point in the field considering the data from all recording electrodes simultaneously.



The measure of global field power (GFP) corresponds to the spatial standard deviation, and it quantifies the amount of activity at each time point in the field considering the data from all recording electrodes simultaneously resulting in a reference-independent descriptor of the potential field. We use L1 norm to calculate the GFP peaks.

$$GFP\,(L_1\ Norm) = \frac{\sum_i^K |V_i\ (t)\ - V_{mean}(t)\ |}{K}$$

Where,

K - number of channels

V(t) - EEG state represented by a (K x 1) vector

<u>EEG Microstates - Algorithm</u>

[In most scenarios, number of EEG recording channels (N_s) >> number of microstates (N_μ) as with EEG microstates we are trying to represent an EEG time-series with the least possible number of states.]

Let us consider a model with just 2 EEG recording channels ($N_s = 2$). Thus we have a 2-D plane that represents all possible electric potential values, and a single point corresponds to an instantaneous measurement.

For a classification model to be effective, it needs to have a minimum of 2 clusters, thus in this scenario we have 2 microstates. ($N_\mu = 2$)

Now, we initialize the two brain microstates as random vectors which are at unit distance from the origin. All points lying on the line going from the origin toward the microstate belong to the same microstate.

In mathematical terms the final aim for microstate analysis is,

$$V_t = \sum_{k=1}^{N_\mu} a_{kt}\Gamma_k$$

Where,

V_t = (N_s x 1) vector consisting of the scalp electric potential measurements

Γ_k = is the normalised N_s x 1 vector representing the k^{th} microstate

a_{kt} = the k^{th} microstate intensity at time instant t.

Moreover, every EEG state can only be represented by just one microstate i.e. in order to allow for non-overlapping microstates at each time instant t, all a_{kt} must be zero except for one. Therefore, at each time instant, the summation reduces to a single nonzero term, corresponding to a single active microstate.

Now, the above mentioned equation will not be able fit perfectly especially when ($N_s >> N_\mu$) or when the EEG time series is long ie. we have many EEG states. Thus we add a term for error.

$$V_t = \sum_{k=1}^{N_\mu} a_{kt}\Gamma_k + E_t$$

The goal will be find microstates such that

1) Every EEG state will belong to **only** one microstate
2) The total error in the system will be minimum. The orthogonal squared distance between each measurement vector and microstate is computed. The higher the distance the higher the error.

$$d_{kt}^2 = V_t'\,.\,V_t - (V_t\,.\,\Gamma_k)$$

Let us assume we now have 32 electrodes and choose to classify them into 4 microstates.

1) We have a 32 dimension space, with every point a possible EEG state. Dimension of the EEG state here is 32x1.
2) We randomly construct random vectors in the above mentioned space of unit length. The dimension of the microstate is 32x1.
3) We now use k-means algorithm which gives the 4 most optimal microstates; and classifies all EEG states into one of those microstates with the orthogonal squared distance (error) less than the threshold error which was set by the user.

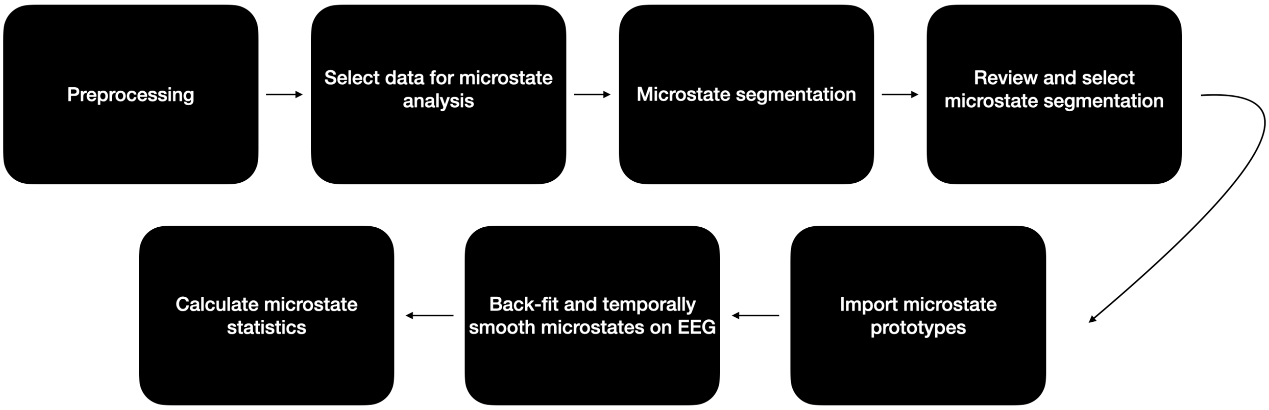
<u>EEG Microstates - Toolbox tutorial implemented</u>

The EEG Microstate analysis toolbox given in MATLAB is used in this tutorial. The dataset associated with the tutorial had the following characteristics

1) Resting state EEG data
2) 4 Test subjects; age group 25-44
3) Pre-Processed and filtered with a high-pass filter of 1 Hz and a lowpass filter of 30 Hz

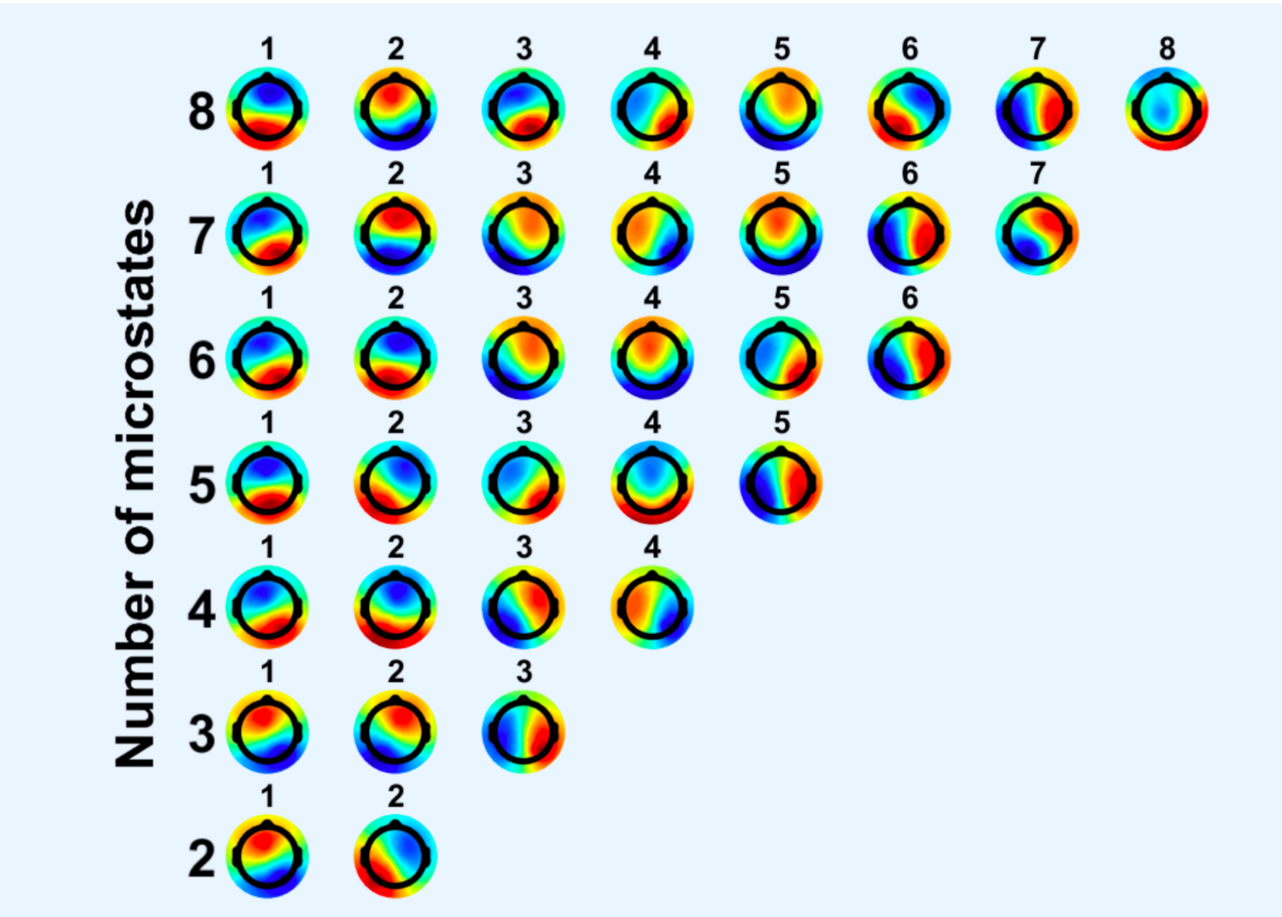
GOAL: parse EEG maps into microstate prototypes and re-express the spatio-temporal characteristics of the EEG time series by means of the microstate prototypes

TUTORIAL PIPELINE



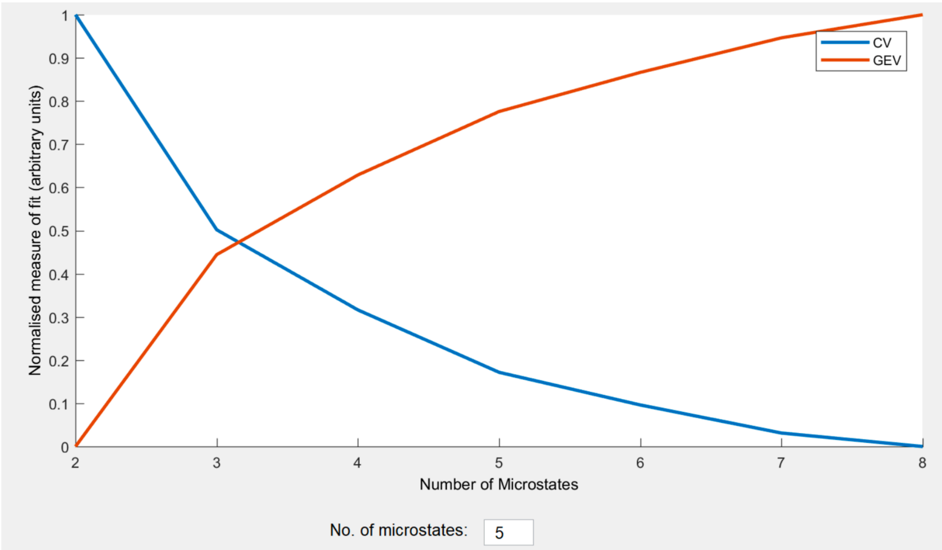
Result: Microstate segmentation

We now perform microstate segmentation for k = [2,8] where k is the number of microstates.



Choosing the number of Microstates

1) Experimentally, k=4 is the most common metric used in most of the resting state EEG scenarios
2) We use two methodologies to determine the value of k. GEV (Global explained variance) and CV (Cross-validation criterion).
3) GEV - is a measure of how similar each EEG sample is to the microstate prototype it as been assigned to. GEV uses GFP which quantifies the amount of activity at each time point in the field considering the data from all recording electrodes simultaneously. Thus the aim is to maximise GEV.
4) CV - This measure is related to the residual noise, ε, and the goal is therefore to obtain a low value.



We see that GEV is maximum for 8 microstates and CV was minimum for 8 microstates, which is theoretically expected. Therefore for computationally efficiency we have stop when adding another cluster does not bring a significant benefit. Here we get the number of microstates = 5.

The final blocks of the pipeline

Back-fit microstates on EEG - assigning a microstate labels to EEG states based on which microstate prototype they are most topographically similar with. This similarity can be measured using global map dissimilarity (GMD).

Temporally smooth microstate labels - Noise is quite significant whilst using resting state data and hence we apply label smoothing.

Calculate microstate statistics - Using the toolbox we can calculate different statistics like average GFP in all time frames of the same microstate class, the occurrence of each microstate class per second, the average duration, the percentage coverage of each microstate class, the

Global Explained Variance and spatial correlation (i.e. on average how much variance in the EEG data is explained by the best fitting microstate prototype) and the transition probabilities between microstate classes.