# MACHINE LEARNING APPROACH TO PREDICT SUCCESSFUL MEMORY ENCODING

Akshay Arora & Bradley Lega
UTSouthwestern Medical Center

# Restoring Active Memory (RAM) Project

- Four year $22.5 million grant from DARPA under BRAIN init
- Collaboration between 13 sites

- University of Pennsylvania
- **UTSW**
- Thomas Jefferson University
- Dartmouth
- Emory
- NINDS
- Mayo
- Columbia
- Boston University
- Ohio State University
- Medtronic
- Lawrence Livermore National Labs
- Blackrock Microsystems

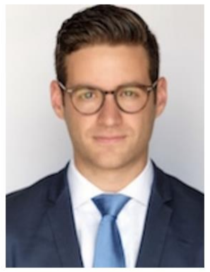
- 190 total subjects (44 from UTSW)
    - Epilepsy patients with intracranial electrode implants

UT SOUTHWESTERN MEDICAL CENTER

## Principal Investigator

**Dr. Bradley C Lega**

## Lab Members



Jimmy Germi

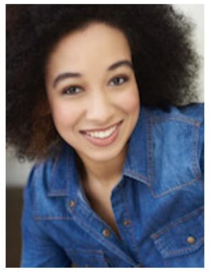
Clinical Data Specialist
Biological Basis of Behavior
University of Pennsylvania, 2015

Jui-Jui Lin

Clinical Data Specialist
Electrical Engineering (BE), Southern Methodist University, 2015

Akshay Arora

SMU master's student in Electrical Engineering



Alexis Smith

Clinical Data Specialist, IRB Liaison
MS Candidate, Applied Cognition and Neuroscience, UT Dallas

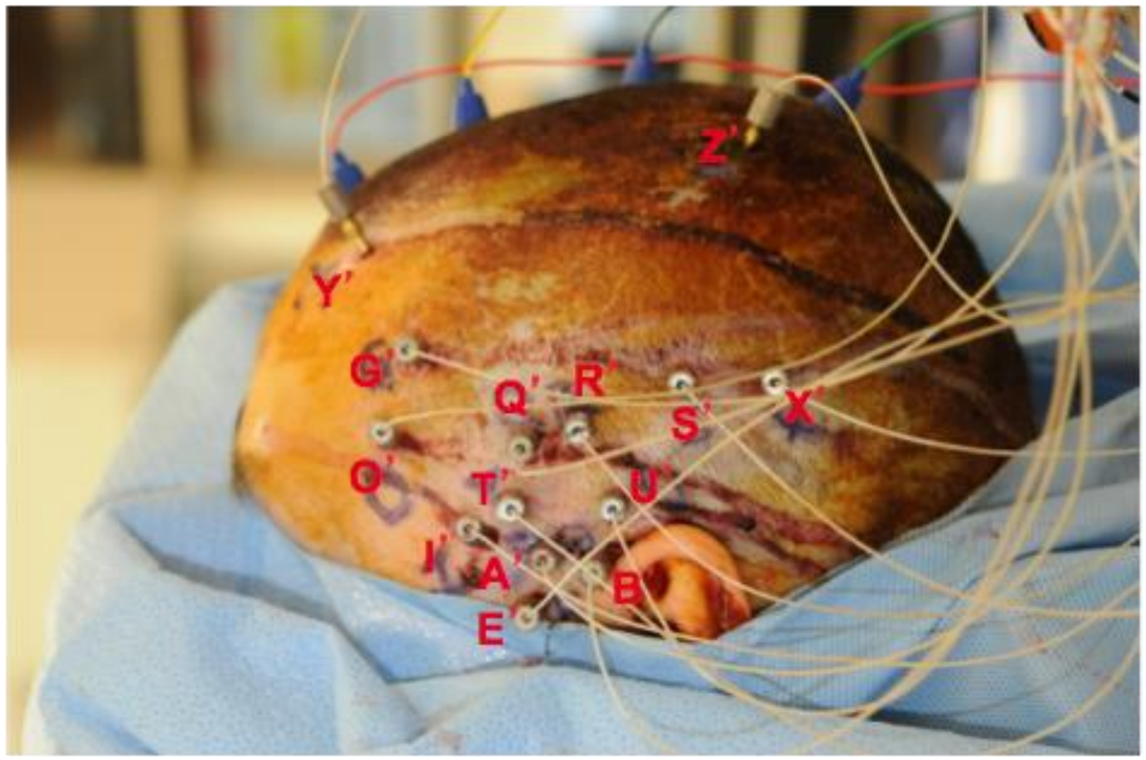
Xiao Xu

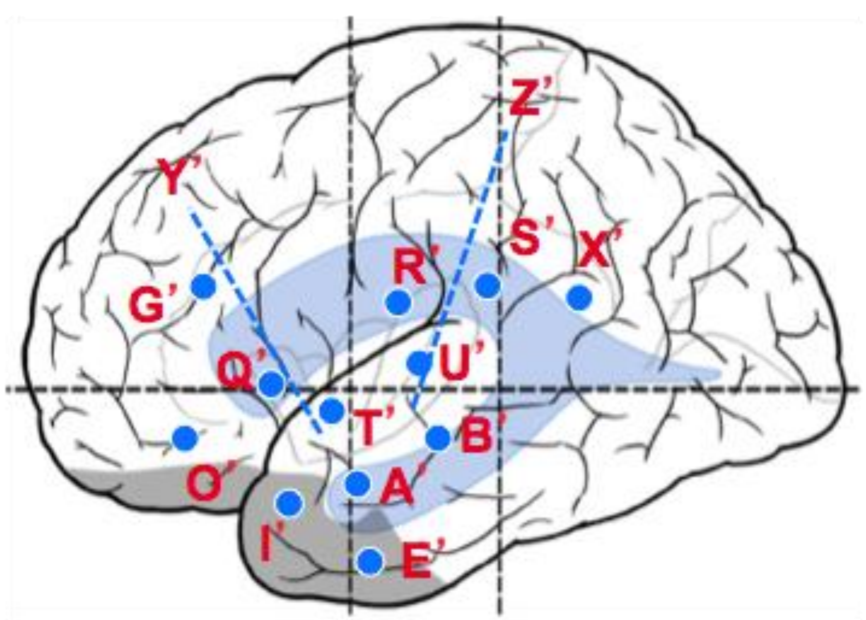
Data Analyst
ME, Electrical Engineering, SMU, 2015
Electrical Enginger, Ph.D. Candidate, SMU, 2015

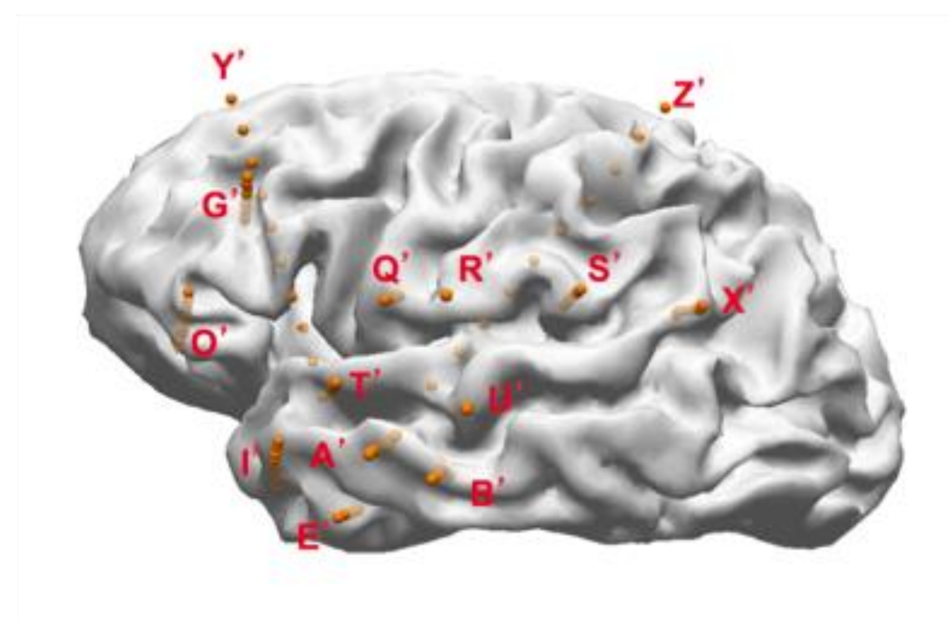
Constance Kemp

Program Coordinator

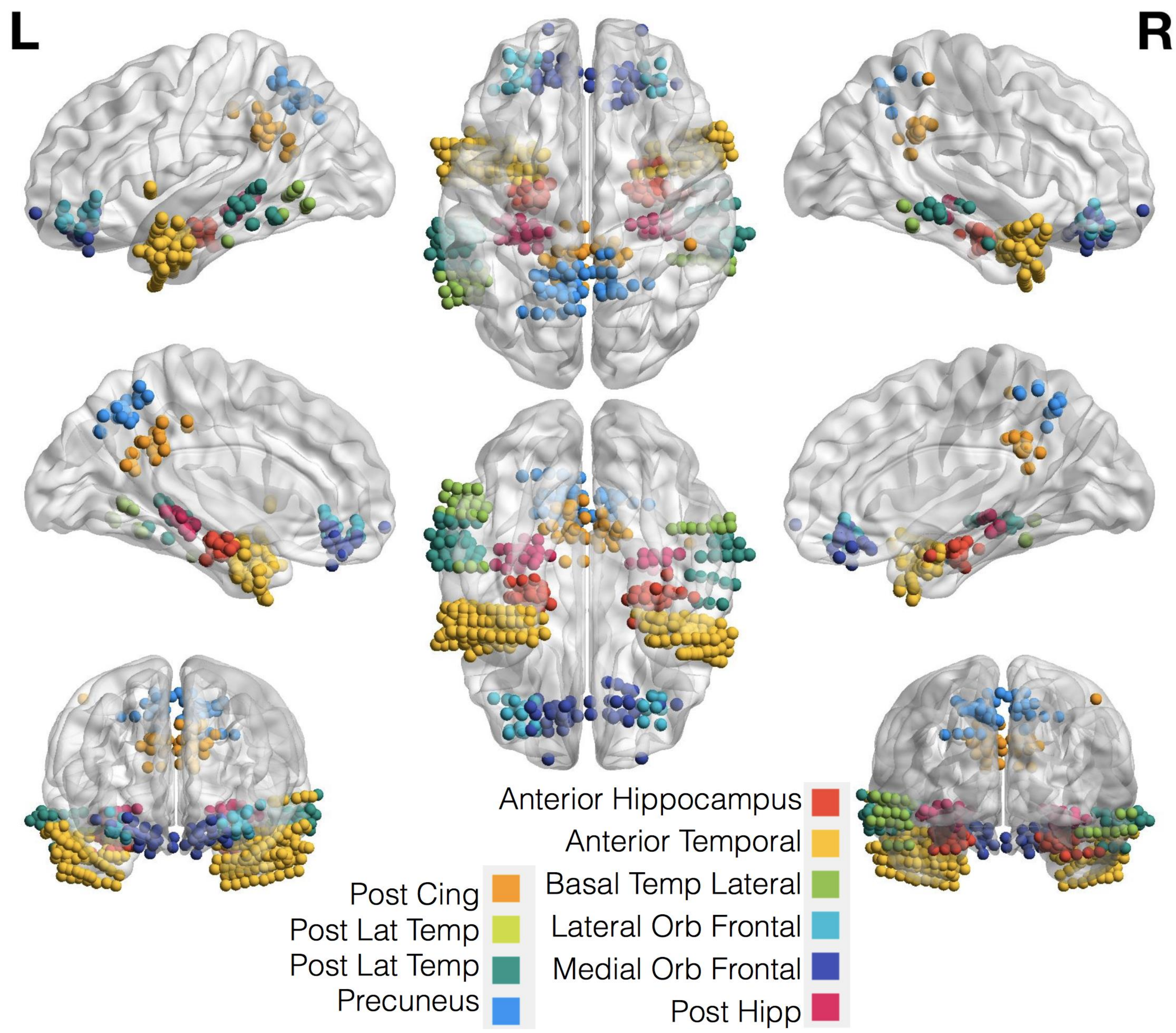
# Stereo EEG Intracranial Depth Recording



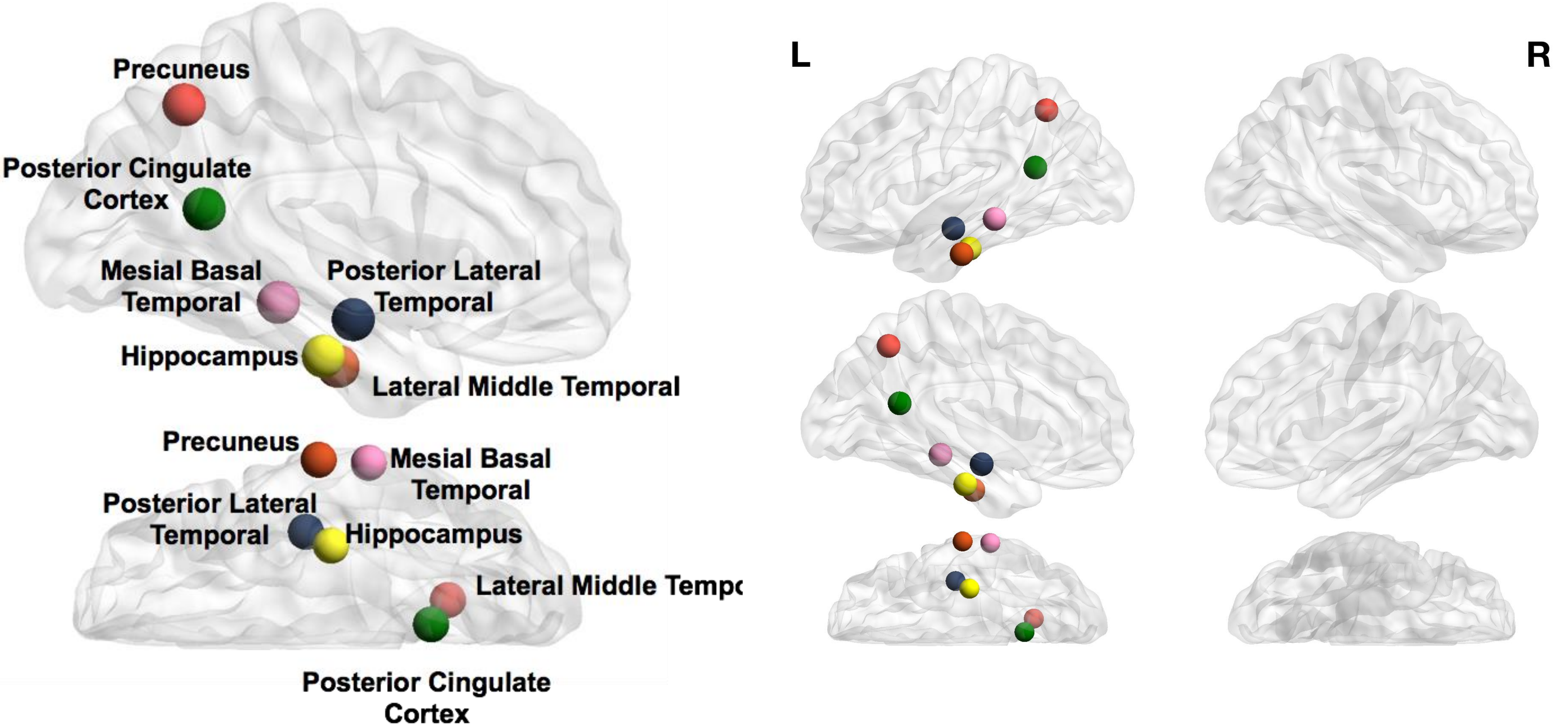



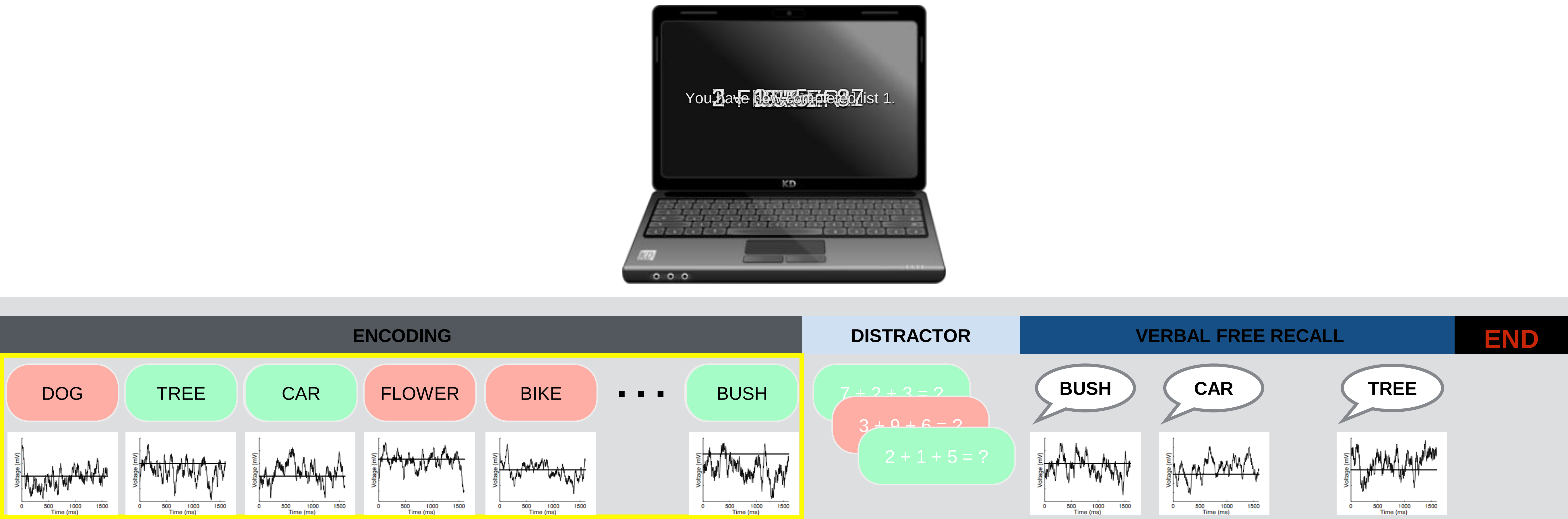
# Electrode Localization



# Regions of Interest

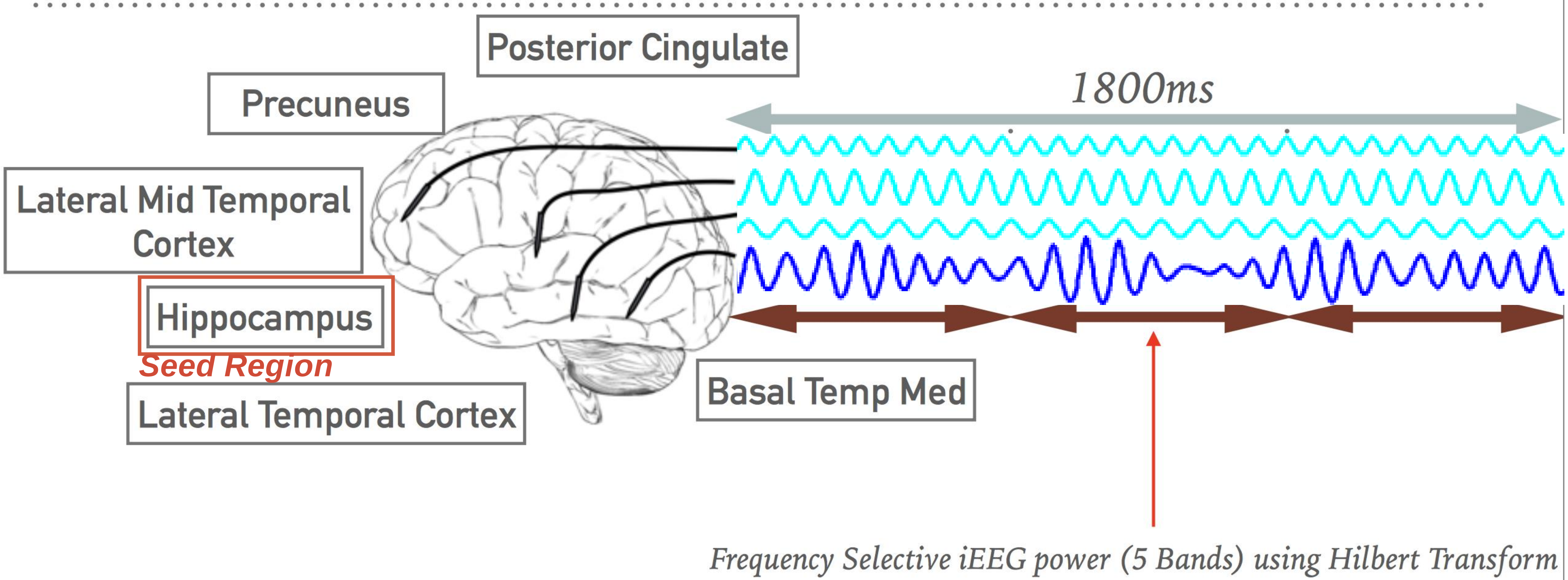



***Seed Region : Hippocampus***




Able to compare neural activation patterns during the **encoding** of words that are later remembered with patterns during **encoding** of words that are not remembered

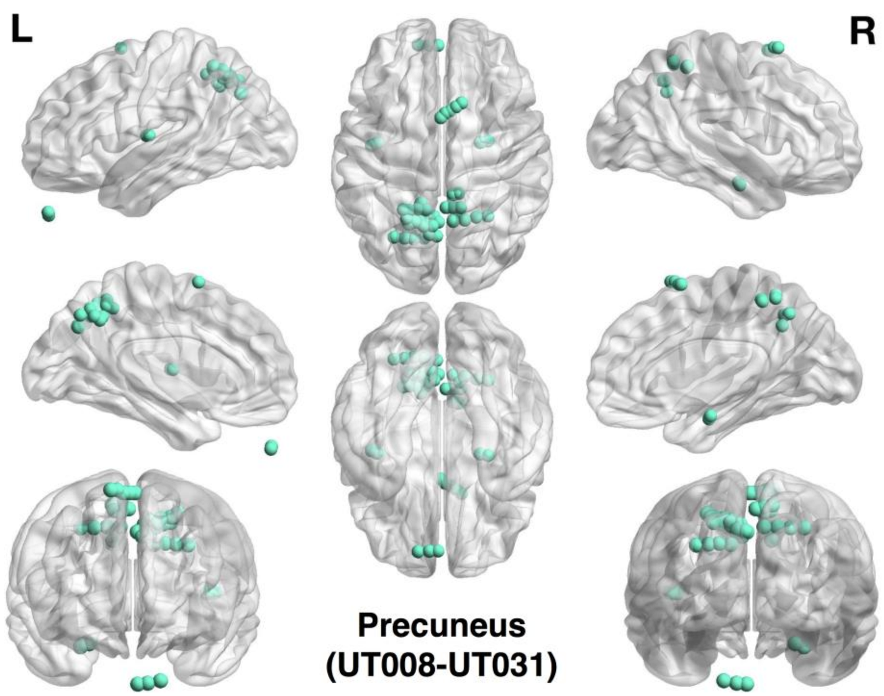
# DATA ACQUISITION AND PREPROCESSING : POWER DATA



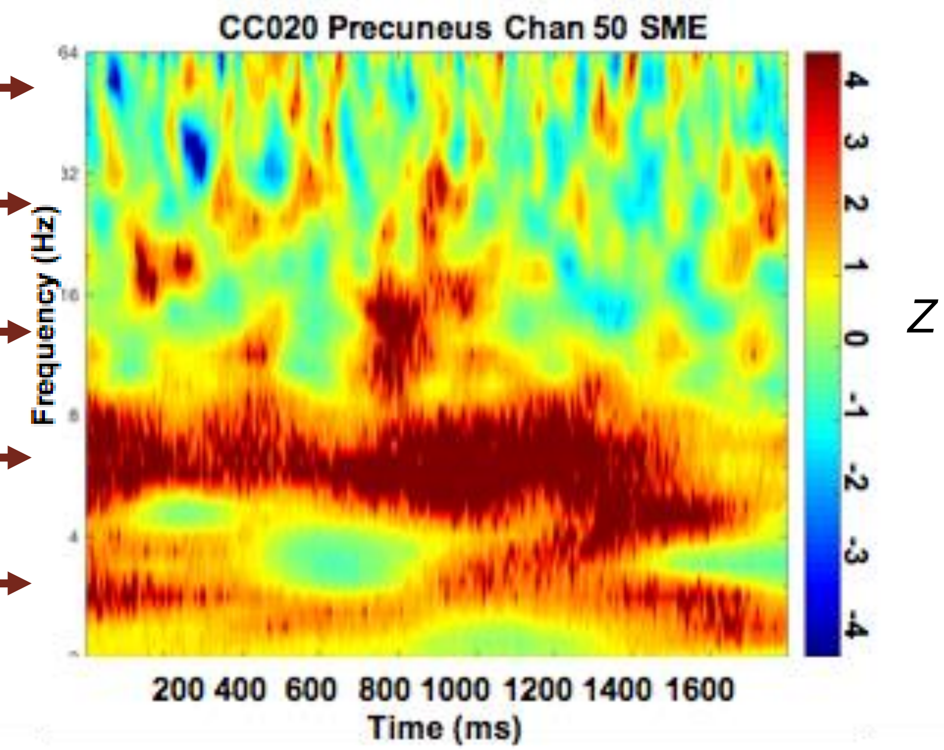

Delta: 2-4Hz
Theta: 4-9Hz
Alpha : 9-16Hz
Beta :16-32Hz
Gamma: 32-55Hz

Data : Events x Frequency(5 Bands) x Time(3 Windows) x 6 Regions )

# Intracranial EEG Power Spectrogram



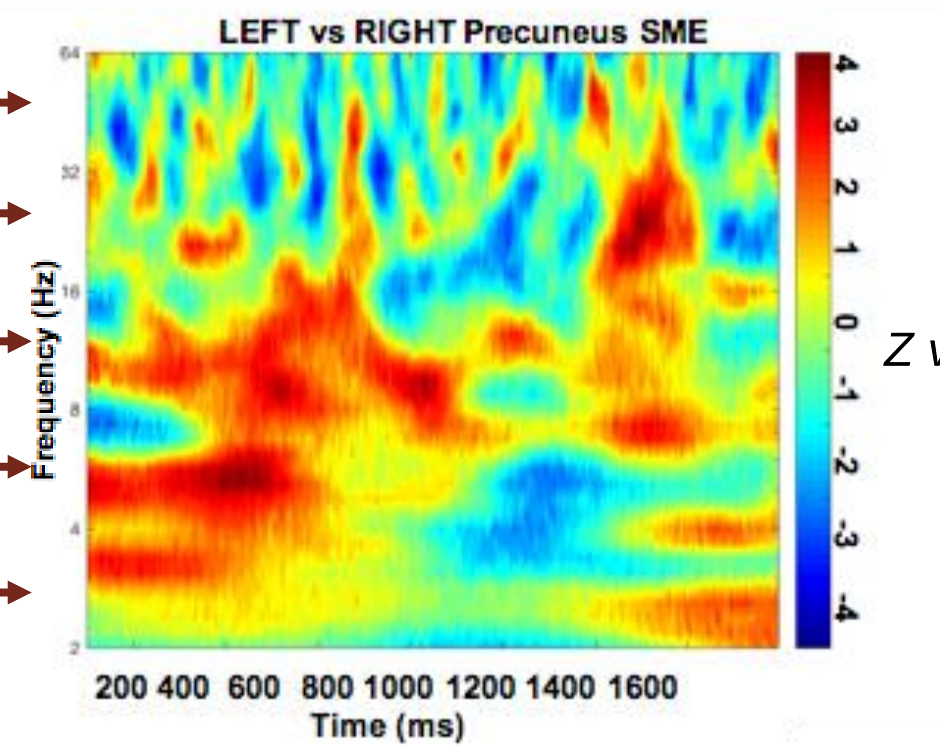

*Gamma(25-55Hz)*
*Beta (16-25 Hz)*
*Alpha (9-16Hz)*
*Theta (5-9Hz)*
*Delta (2-5Hz)*




*Z values*

**Figure 2a : Single Channel Precuneus SME**

*Gamma(25-55Hz)*
*Beta (16-25 Hz)*
*Alpha (9-16Hz)*
*Theta (5-9Hz)*
*Delta (2-5Hz)*



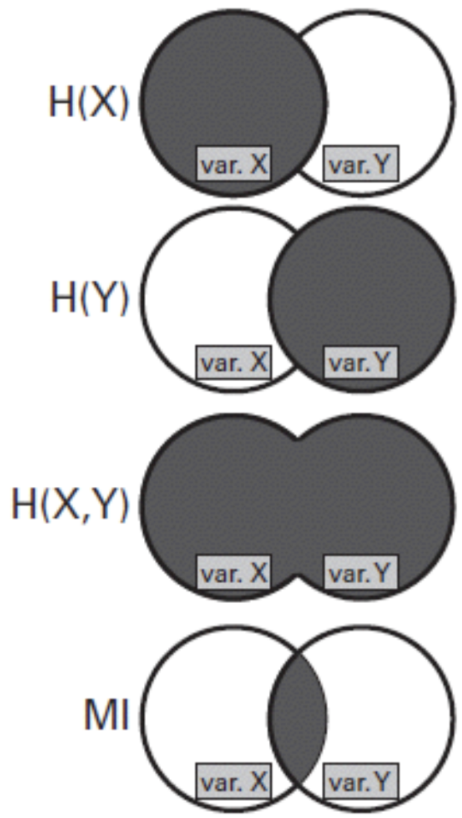

*Z values*

**Figure 2b: Left vs Right Precuneus SME**

# DATA SET : MUTUAL INFORMATION DATA (ALL SIX REGIONS)

- *The MI values are **Z-normalized** and the are averaged across channel pairs.*
- *The max positive and **negative** lags across each event is chosen to be the predictor for 'recall' and 'non-recall' for all 6 regions of interest.*

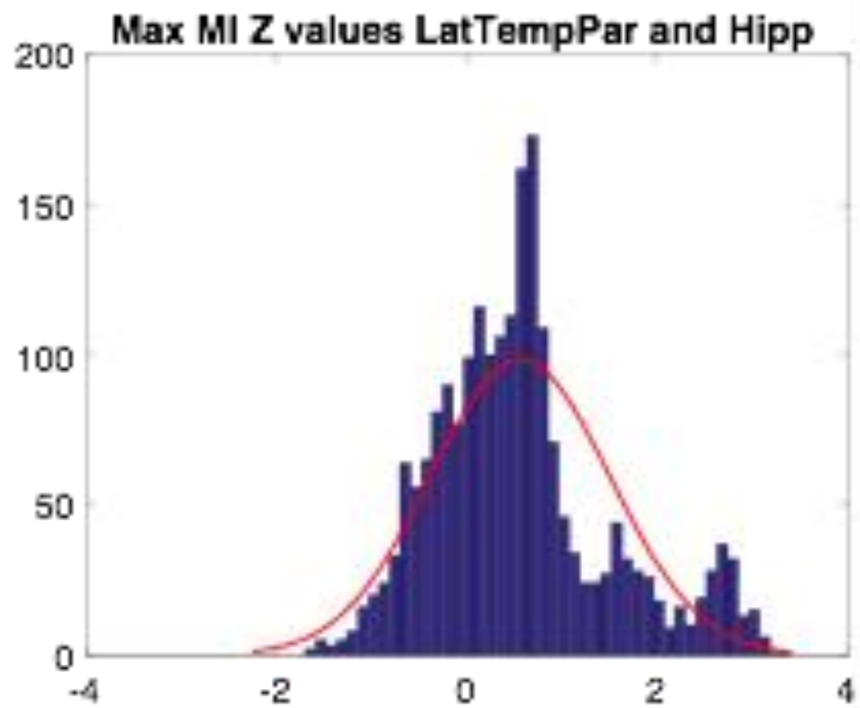
$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$




$$MI(X,Y) = H(X) + H(Y) - H(X,Y)$$

*Data : Events x ChannelPairs(Seed: HIPP) x TimeLags ( -200ms to 200ms)*

# Mutual Information (Lagged)



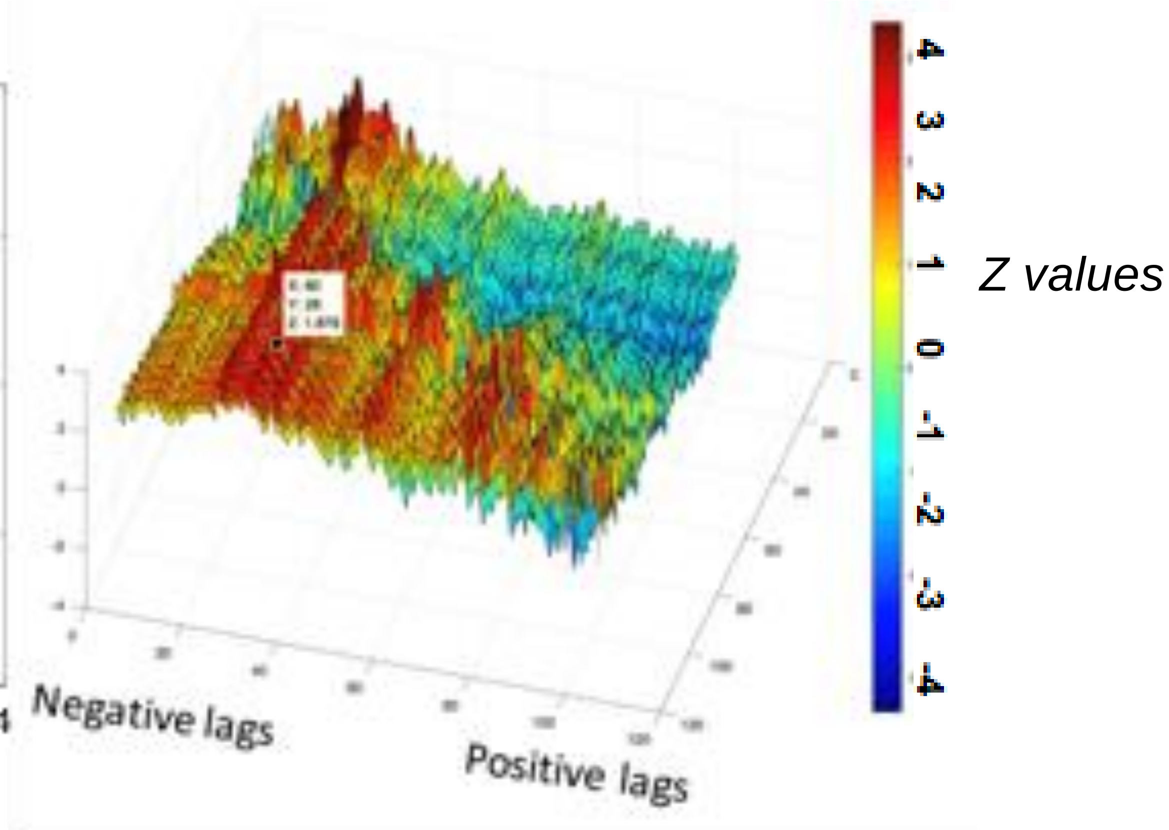




Figure 3a : Max MI Z Val Distribution

- Max MI Z Values are Normally Distributed.
- Precuneus leads the Hippocampus by 112ms.

# Machine Learning Data

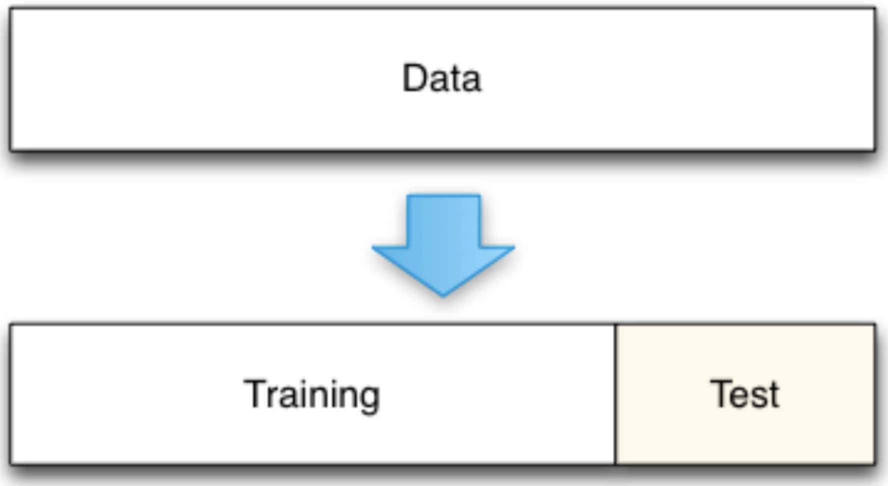
**Features**

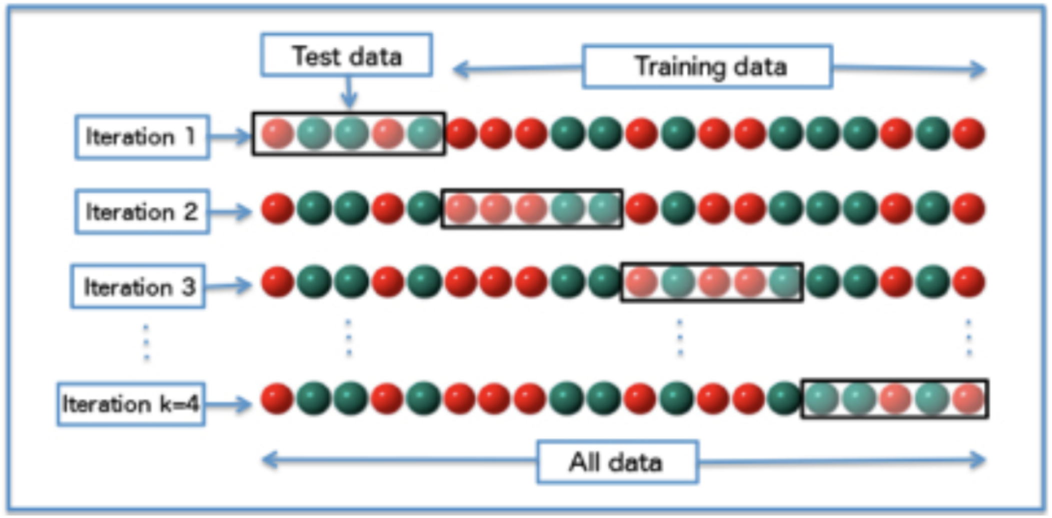
**Events (Trials)**

| Power Region 1 | Power Region 2 | Power Region 3 | Power Region 4 | Power Region 5 | Power Region 6 | MI Region 6 | MI Region 6 | MI Region 6 | MI Region 6 | MI Region 6 | MI Region 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | |
| Recall | | | | | | | | | | | |
| Recall | | | | | | | | | | | |
| Recall | | | | | | | | | | | |
| NonRecall | | | | | | | | | | | |
| NonRecall | | | | | | | | | | | |
| NonRecall | | | | | | | | | | | |
| … | | | | | | | | | | | |

Data : Events x Frequency(5 Bands) x Time(3 Windows) x 6 Regions )

# CROSS VALIDATION TECHNIQUES



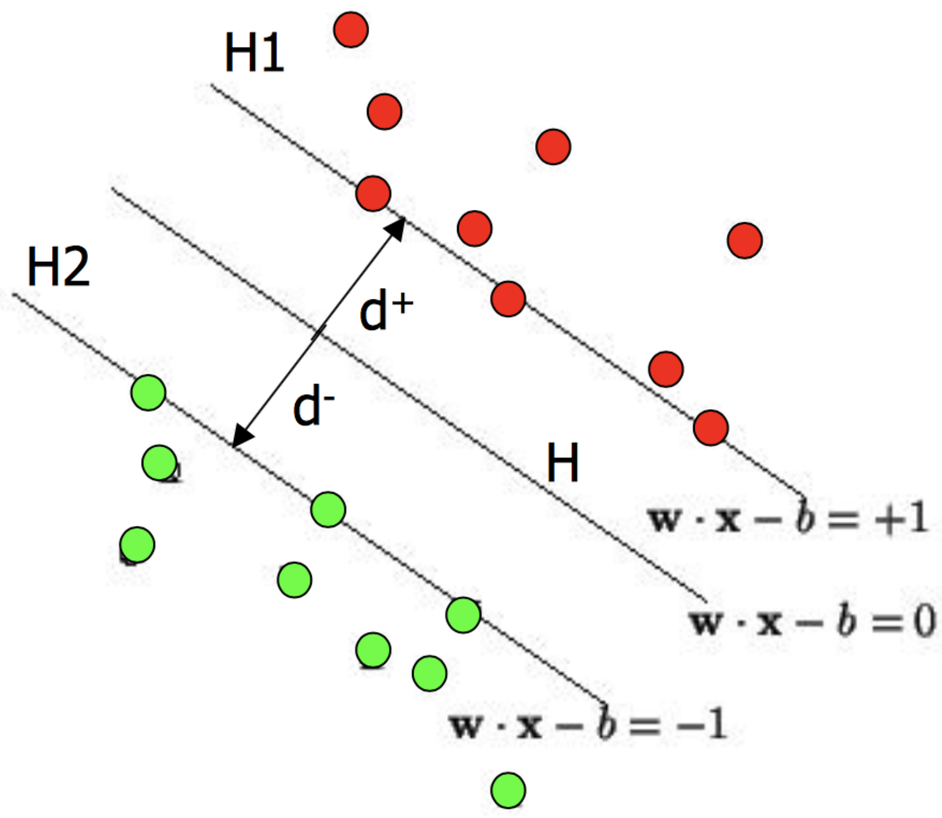

- *HoldOut Validation : 70 % : Training , 30 % : Testing*
- *k-Fold CrossValidation : Divide Dataset into k-partitions*

# Support Vector Machines

Define the hyperplane H such that:

$\mathbf{x}_i \bullet \mathbf{w}+b \geq +1$ when $y_i = +1$

$\mathbf{x}_i \bullet \mathbf{w}+b \leq -1$ when $y_i = -1$




H1 and H2 are the planes:

H1: $\mathbf{x}_i \bullet \mathbf{w}+b = +1$

H2: $\mathbf{x}_i \bullet \mathbf{w}+b = -1$

The points on the planes H1 and H2 are the Support Vectors

d+ = the shortest distance to the closest positive point

d- = the shortest distance to the closest negative point

The <u>margin</u> of a separating hyperplane is $d^+ + d^-$.

# Support Vector Machines

## We now must solve a <u>quadratic</u> programming problem

- Problem is: <u>minimize</u> $||\mathbf{w}||$, **s.t.** discrimination boundary is obeyed, i.e., min $f(x)$ s.t. $g(x)=0$, where

  $f$: $\frac{1}{2}\,||w||^2$ and

  $g$: $\mathbf{y_i(x_i \bullet w)\text{-}b = 1}$ **or** $\mathbf{[y_i(x_i \bullet w)\text{-}b] - 1 = 0}$

This is a **<u>constrained optimization problem</u>**

Solved by Lagrangian multipler method

# Support Vector Machines

## In general

Gradient max of $f$

constraint condition $g$

$L(x,\alpha) = f(x) + \sum_i \alpha_i g_i(x)$ a function of $n+m$ variables

$n$ for the $x's$, $m$ for the $\alpha$. Differentiating gives $n+m$ equations, each set to 0. The $n$ eqns differentiated wrt each $x_i$ give the gradient conditions; the $m$ eqns differentiated wrt each $\alpha_i$ recover the constraints $g_i$

In our case, $f(x)$: $\frac{1}{2}\|\mathbf{w}\|^2$ ; $g(x)$: $y_i(\mathbf{w}.x_i + b) - 1 = 0$ so Lagrangian is

$$L = \tfrac{1}{2}\|\mathbf{w}\|^2 - \Sigma\alpha_i[y_i(\mathbf{w}.x_i + b) - 1]$$

# Support Vector Machines

## Lagrangian Formulation

- In the SVM problem the Lagrangian is

$$L_P \equiv \tfrac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l} \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{w} + b) + \sum_{i=1}^{l} \alpha_i$$

$$\alpha_i \geq 0, \forall i$$

- From the derivatives = 0 we get

$$\mathbf{w} = \sum_{i=1}^{l} \alpha_i y_i \mathbf{x}_i, \sum_{i=1}^{l} \alpha_i y_i = 0$$

# Logistic Regression

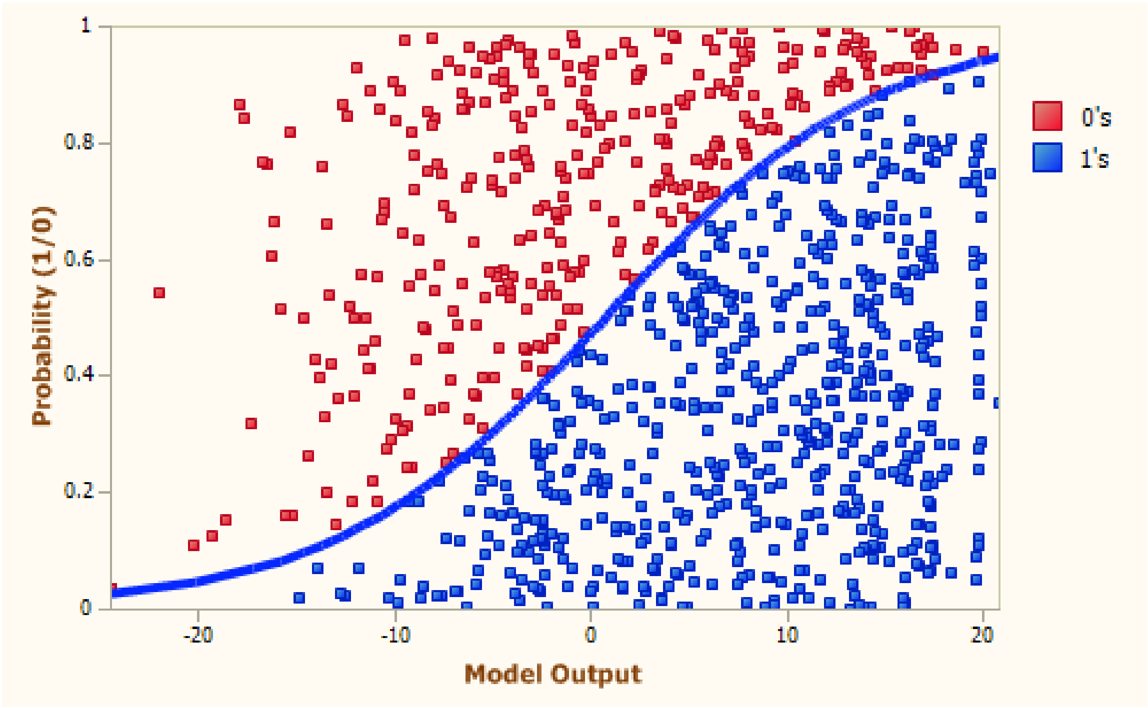



**Figure: Sigmoid function, or logistic function**

- Logistic regression is a regression model where there is a binary dependent variable (DV) where it can take only two values, such as pass/fail, win/lose, alive/dead or healthy/sick.
- Its activation function is expressed as follows :

$$S(t) = \frac{1}{1+e^{-t}}$$

Where, $t = \beta_0 + \beta_1 x$

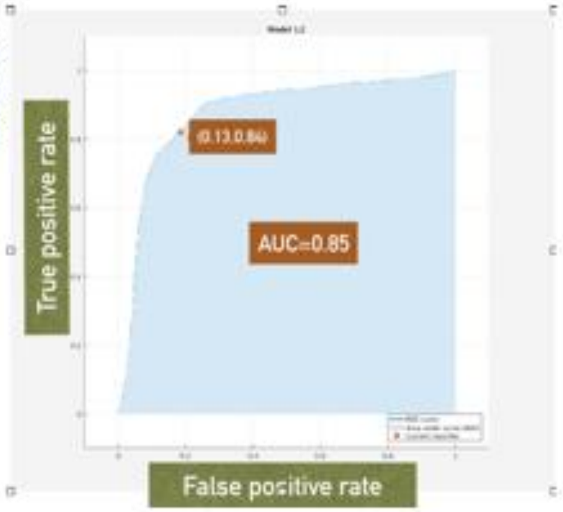
and

$$F(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$$

Binary Classifier : Probabilistic Model

## SVM Classifier : (quadratic) Combined Subjects

- The ROC shows that the SVM model predicts recall performance with an of AUC =0 .85
- Recall Accuracy = 84 %

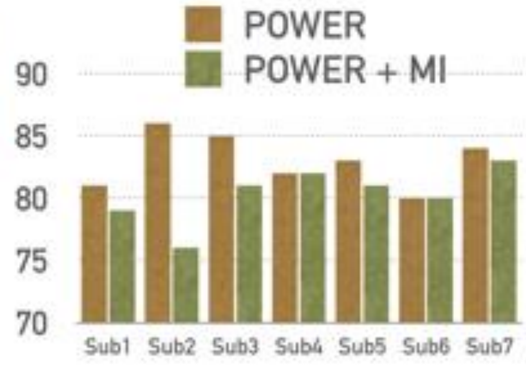



## Logistic Regression Classifier : Combined Subjects

- The ROC shows that the logistic Regression model predicts recall performance with an of AUC =0 .75
- Recall Accuracy = 0.75 %


## SVM Classifier : (quadratic) Individual Subjects

- The Bar plots shows that the SVM model predicts recall performance with an average of 82 % accuracy across all subjects.
- Power gives similar prediction results compared to than choosing Power + MI




## Logistic Regression Classifier : Individual Subjects

- The Bar plots shows that the logistic model predicts recall performance with an average of 76 % accuracy across all subjects.
- Power gives similar prediction results compared to than choosing Power + MI

# Cross Validation Error

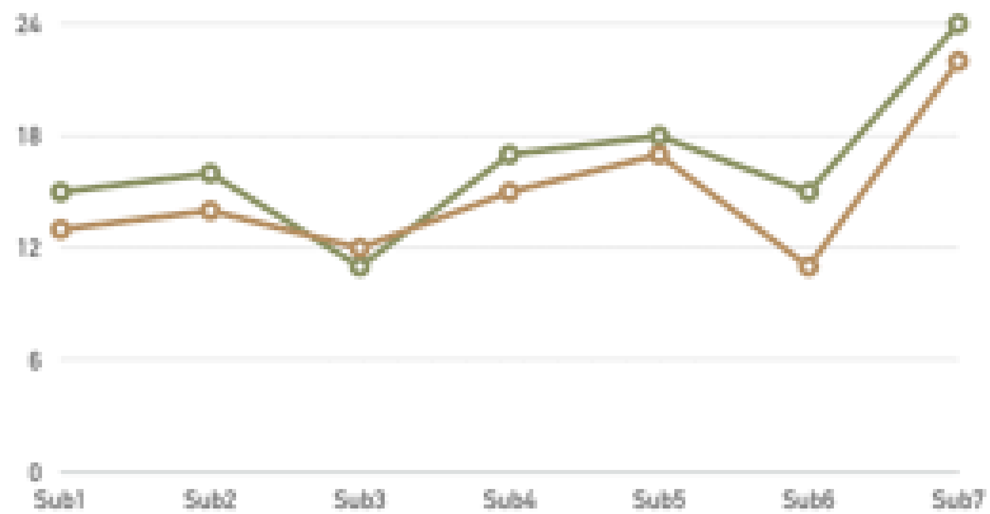



**Figure a :K-Fold Loss (Support Vector Machines Classifier)**

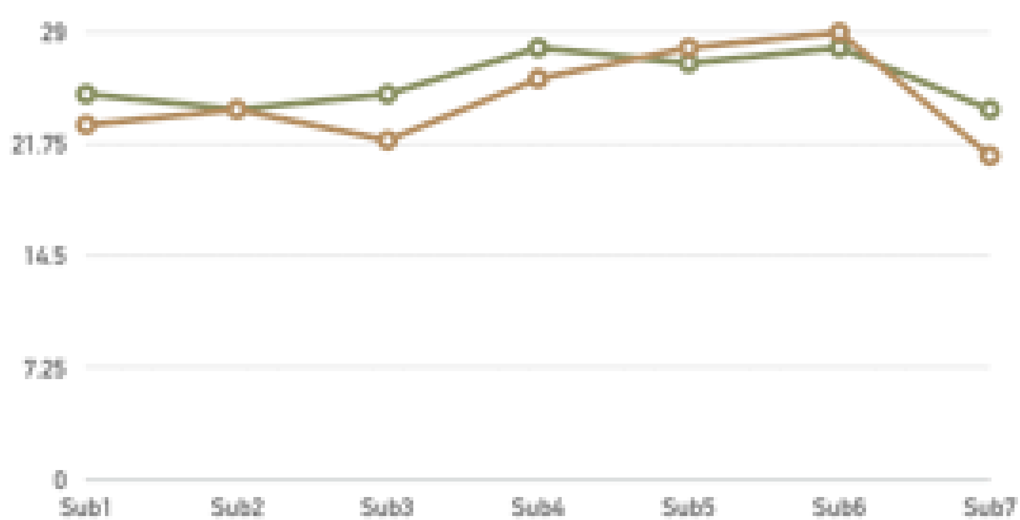
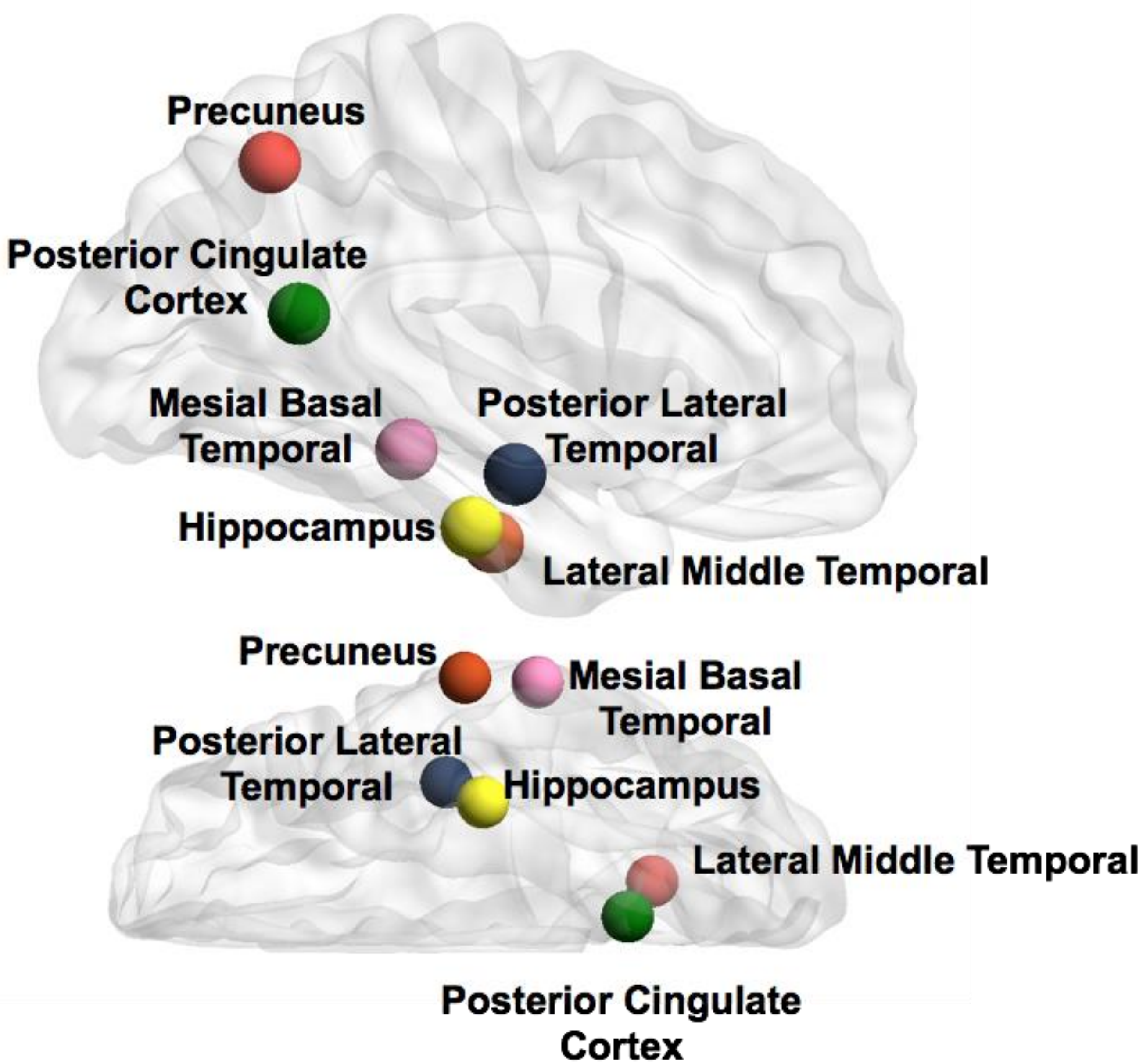



**Figure b :K-Fold Loss (Logistic Regression Classifier)**

UT SOUTHWESTERN MEDICAL CENTER




## Classification Summary (Leave one Region Out)

Table 1 :Classification Summary (Leave one Region Out)

| Region Excluded Out | AUC | Recall % |
|---|---|---|
| Basal Temp Mesial | 0.85 | 83 % |
| Lateral Temporal Cortex | 0.82 | 82% |
| Lateral Middle Temporal Cortex | 0.83 | 81% |
| Precuneus | 0.79 | 78% |
| Posterior Cingulate | 0.79 | 78% |
| Hippocampus | 0.75 | 73% |

- From this Table above, it can be observed that the prediction accuracy and classifier AUC value reduces when we a generate a classifier without "Hippocampus" Region.

# Future Work

- **Estimation of important predictors from iEEG Power and Mutual Information to improve the classifier performance.**

- **Adding more Connectivity based predictors to the existing models**

**Thank You !**