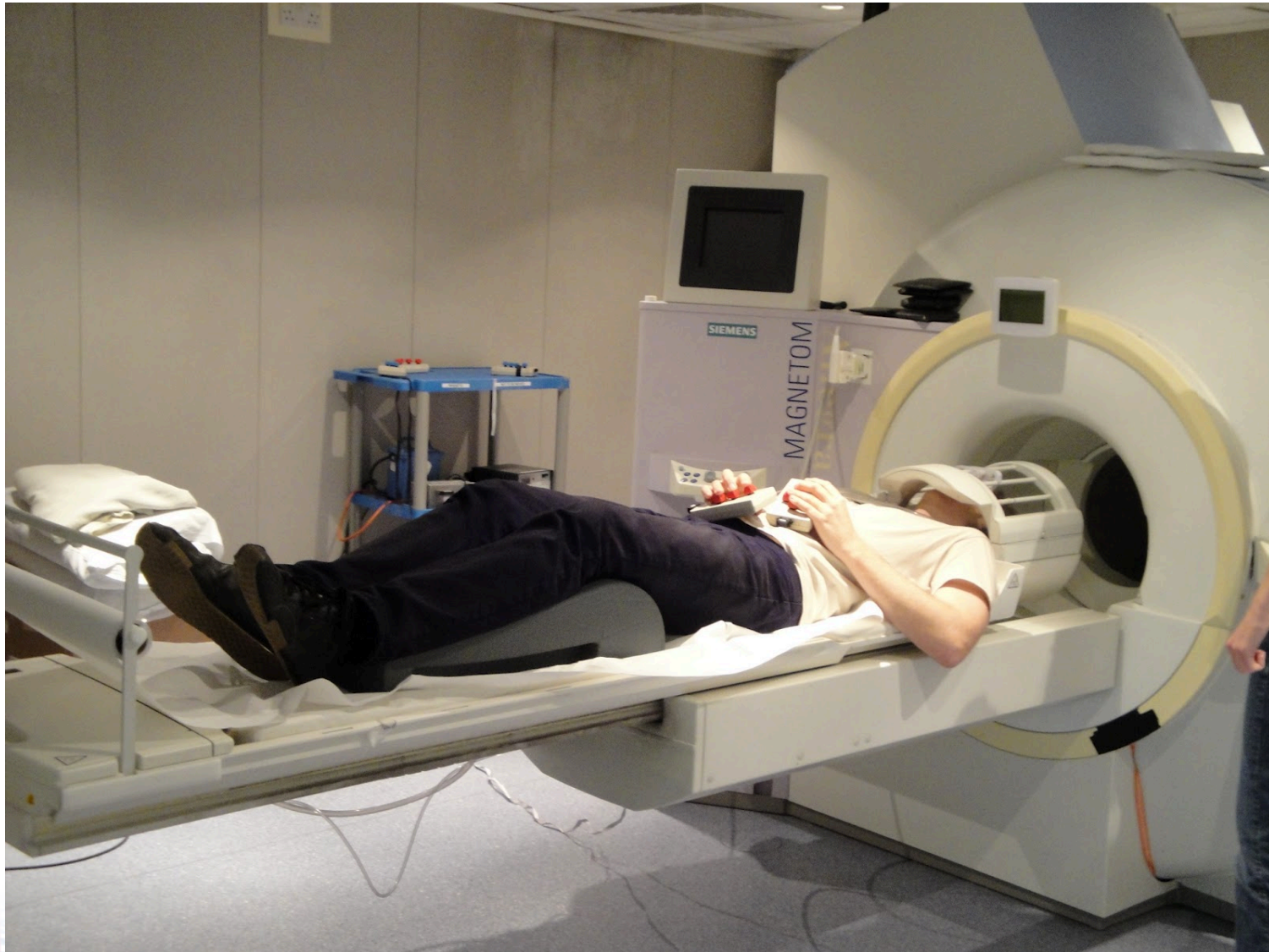# Information Gain in Clustering of FMRI Data
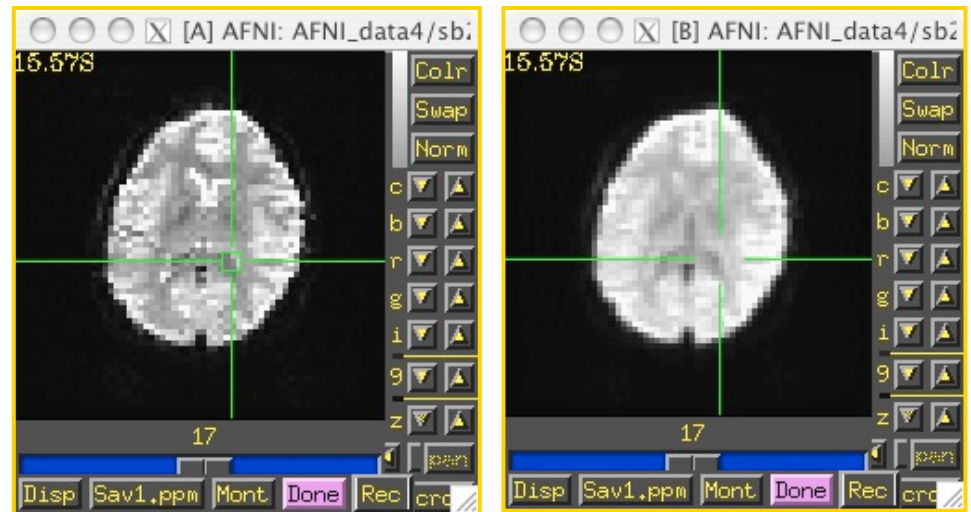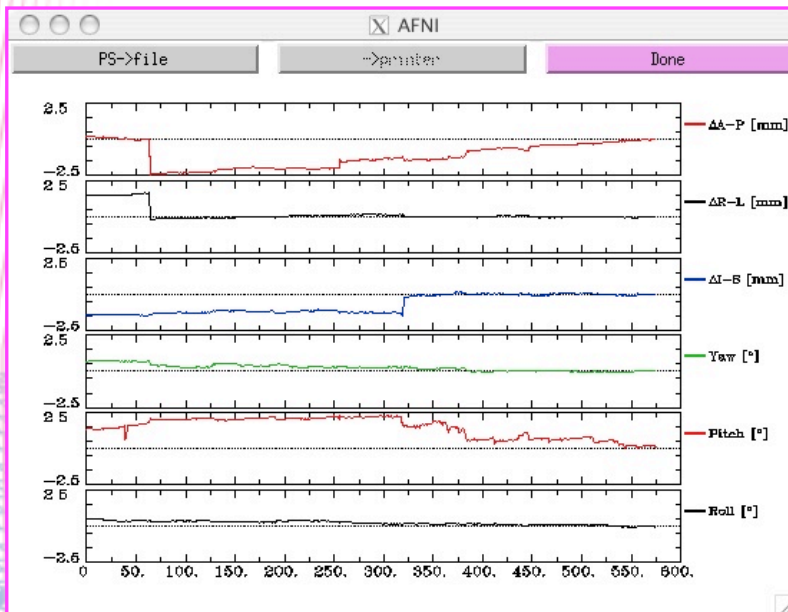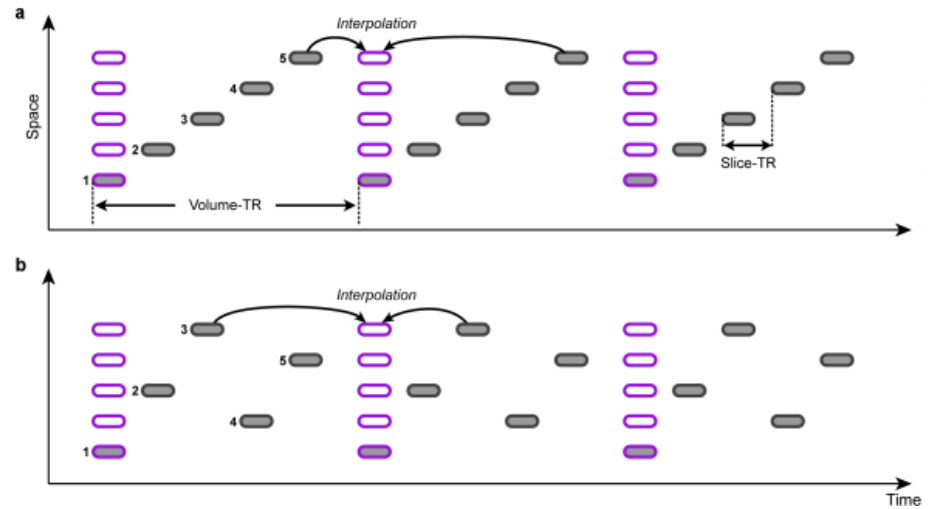
Rafael Bravo

FMRI Scanner

Data Processing

- Time Shift Interpolation

- Motion Correction

- Smoothing

Data Analysis

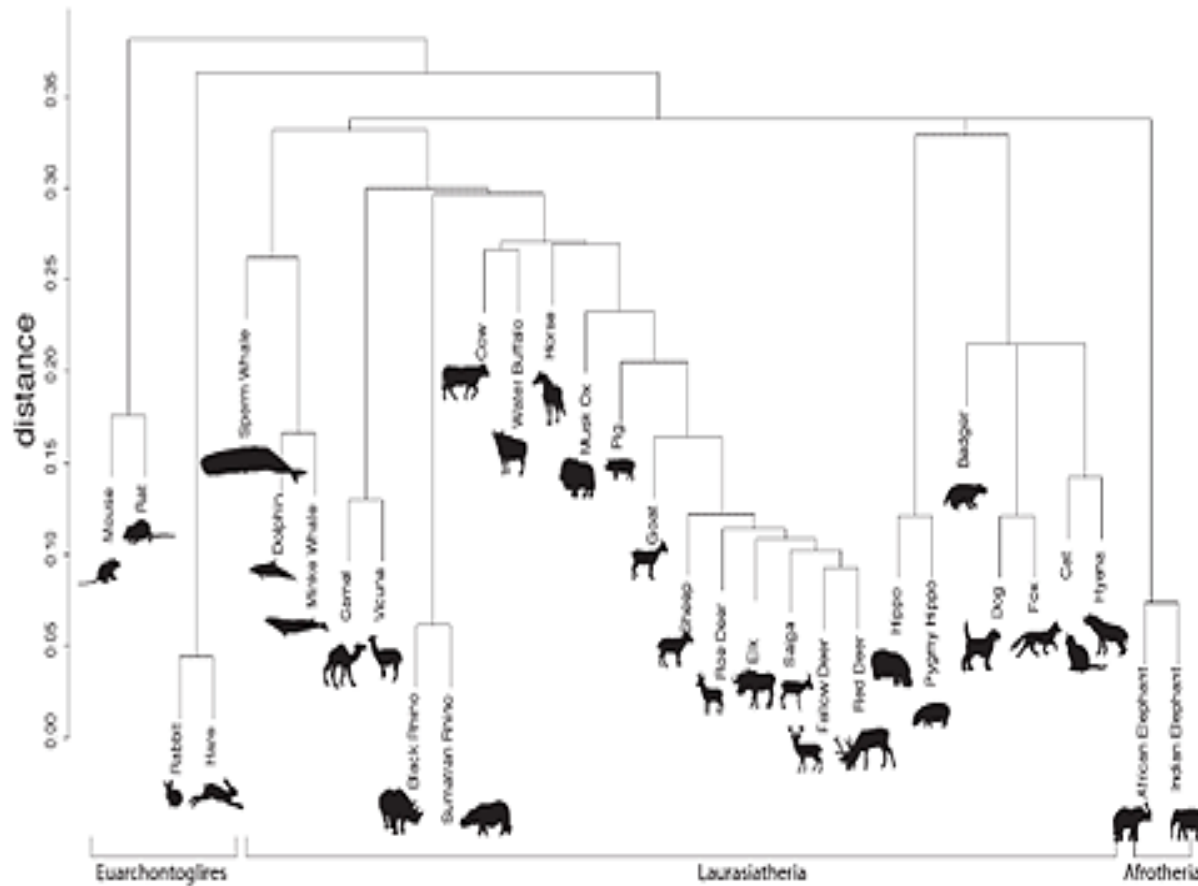Done on the individual voxel level with smoothing

Problems:

Low Signal to Noise Ratio

Large number of simultaneous hypotheses being tested, diminishing significance of results.
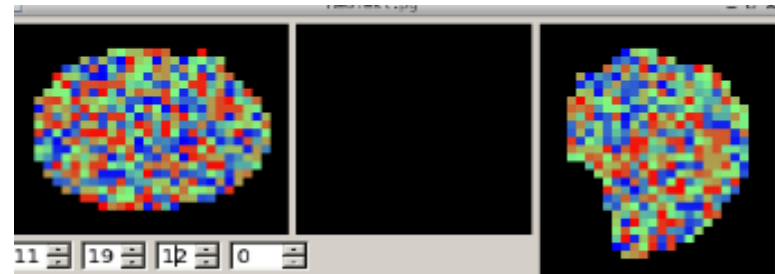
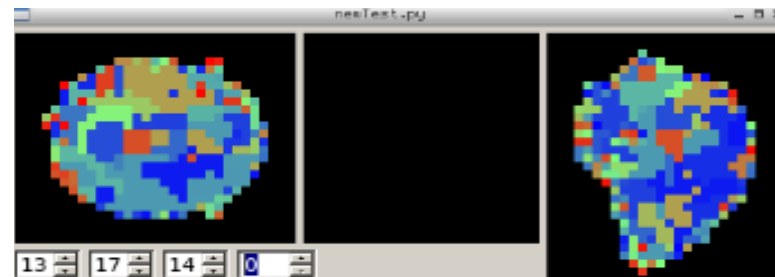Clustering deals with these issues.

Dendrogram Clustering

FMRI Heuristic:
Correlation Coefficient
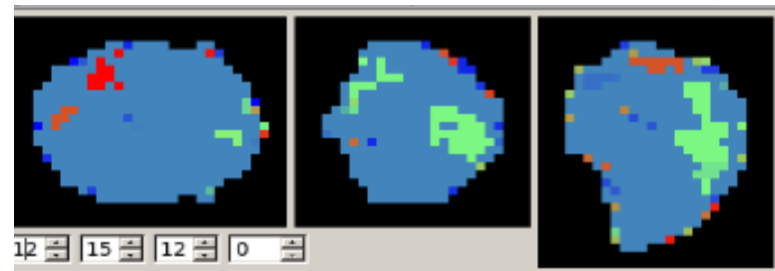
Clustering Results

500 iterations: 4945 clusters

3500 iterations: 1945 clusters

5000 iterations: 445 clusters

Information Questions

How much information is lost due to clustering?

Is there an optimal level of clustering that maximizes statistical significance but minimizes feature loss?

Alternative Clustering Heuristics? Like mutual information, or generated HMM similarity?

"Optimal" Clusters

Can measure information loss by looking at statistical complexity of system at each step?

Once found:

Can look for correlation with experiment events

Can treat clusters as HMMs