# Analysis Plan

Table of Contents

[Analysis Plan 1](#_Toc206249114)

[About 1](#_Toc206249115)

[Experiment Overview 1](#_Toc206249116)

[Pre-processing 1](#_Toc206249117)

[Cleaning 1](#_Toc206249118)

[Parcellation 2](#_Toc206249119)

[HMM Model 2](#_Toc206249120)

[Hypotheses 2](#_Toc206249121)

## About

This analysis plan pertains to the fMRI-BOLD data from a movie viewing task for the HMM Video studies. At the time of writing the analysis plan, data acquisition was completed but no statistical analyses were conducted. See GitHub link for timestamps on analysis pipeline (<https://github.com/jxli25/Video_HMM>). Analyses will only take place once this analysis plan has been uploaded, and time stamped on the Open Science Framework. If the researchers decide to conduct additional analysis, this will be stated in any publication (“post-hoc analyses”).

## Data Collection

The experiment involved showing participants a 4 minute video stimulus. Approximately 40 of these participants had a diagnosed psychotic illness, and 40 were controls. fMRI-BOLD scans were conducted on participants during movie viewing. Participants also completed the following rating scales:

• Positive and Negative Symptom Scale (PANSS)

• Hamilton Depression Rating Scale (HDRS)

• Young Mania Rating Scale (YMRS)

• Simpsons-Angus Scale (SAS) for extrapyramidal side effects

• Clinical Global Impression – Severity (CGI-S), a measure of symptom severity

• Social and Occupational Functioning Assessment Scale (SOFAS), a continuous measure of overall functioning

We from the above complete dataset, we will randomly assign 10 experimental and 10 control participants to a hold-out dataset (HOD) used for analysis.

## Pre-processing

### Cleaning

Scans were normalised to the the MNI152NLin6Asym standard space. Head-motion related movement artefacts were removed. A band-pass filter was applied (high pass = 0.01Hz, low pass = 0.15Hz) to filter out large scale frequency drifts and physiological noise. Spatial smoothing was not applied.

### Parcellation

fMRI-BOLD sequences were standardised and parcellated according to the Yeo-17-thick atlas.

## HMM Model

### Training

We will train an HMM Model on our data after subtracting a hold-out dataset of 10 control group and 10 experimental group participants. The settings as per Table 1 will be used for this model.

Table HMM Model input settings for object Options on MATLAB using the HMM-MAR package

|  |  |
| --- | --- |
| Options. \_\_\_\_\_\_\_ | Setting |
| K | 15 |
| covtype | full |
| DirichletDiag | Will use HOD to estimate optimal DirichetDiag out of candidate priors (0.5, 1, 2, 5, 10).  For each candidate\_prior, fit HMM to the training set with the specified Dirichet prior.  For each trained model, compute how well it predicts the hold-out data.   * Predictive log-likelihood; use learned parameters to evaluate likelihood of the hold-out sequence * Free energy; Evaluate the variational free energy on unseen data   Select the DirichletDiag that maximises the predictive log-likelihood on hold-out set. |
| cyc | 300 |
| initrep | 10 |
| initcyc | 10 |

### Outputs, Analyses and Hypotheses

The outputs of interest for this analysis are described in Table 2. Most of it borrows heavily from (Meer et al., 2020).

Prior to statistical testing, appropriate checks will be made to check assumption are fulfilled (e.g. normalcy, variance). If unfulfilled, table will be updated with updated methods.

Table Outputs of interest for analysis from HMM Model and analysis to be performed

|  |  |  |
| --- | --- | --- |
| Output | Analyses | Hypotheses |
| Hidden States (HS) | 1. Hidden State Decoding: Use 16 general terms of the Neurosynth database as per (Meer et al., 2020). Forward associate each Hidden State to the topic maps of these 16 general terms.   For each HS, calculate the voxel-wise Pearson correlation with each of the 16 terms (Chang et al., 2013).   1. Correlate the spatial distribution of each brain state to the topic maps using this Python notebook.(<https://github.com/neurosynth/neurosynth>). 2. For experimental group, calculate FO for each subject in more ambiguous HSs’ (found after the above) within each segment and for whole video. Calculate Pearson’s correlation co-efficient between higher PANSS, HDRS, CGI-S AND SOFAS and higher FO in ambiguous states for segments and whole video. | 1. NA 2. NA 3. Higher PANSS, HDRS, CGI-S AND SOFAS scores are significantly correlated with increased FO in ambiguous states. |
| Average state paths | 1. Calculating sliding window average state paths: For experimental vs control groups, use a sliding window of 3 *(9 for Meer for a 20 min movie clip, our clip is 4 min, 9/5 approx to 2 however 327 scans is divisible by 3)* consecutive BOLD volumes for each segment and identify the most frequently expressed state, as well as the number of participants that expressed each Hidden State at least once. 2. Consistency: Calculate **consistency** as % of participants expressing the main HS for each sliding window. Calculate overall consistency. Then split data into experimental vs control groups to calculate intergroup differences using paired t test. | 1. NA 2. There is a significant difference in intergroup consistency |
| Transition Probability | 1. Frequency: Apply a threshold of 20% to identify the most frequent transitions, and visualise this. 2. Significance: Use **t-test** to test for significant differences between each transition probability (# HNs x # HNs matrix), and visualise this.   Can use Network-Based Statistics Toolbox (<https://www.nitrc.org/projects/nbs/>) for the above. | 1. NA 2. There are significant differences in transition probabilities between experimental and control groups. |
| Fractional Occupancy (FO) | 1. Correlation with video: Pool all data together. Temporally segment data according to video annotations. Calculate FO of each HN for each segment. **Chi-square test of independence** for FO of each HN for each segment, between segments. (# HN x # segments) Calculate **Cramer’s V** for significant associations. 2. Difference between experimental vs control: Take the same segments, split data into experimental and control groups. Calculate FO of each HN for each segment within each group. Conduct **MANOVA** (use **Benjamin-Hoghberg FDR procedure**, alpha = 0.05). If significance found, uses **state-specific test (t-test)** to identify which states contribute to the difference (also apply BH FDR). *Wald test? For non parametric data unlike ANOVA, where numbers are more categorical in nature (R toolbox). Comedy.* 3. Do the same as 9. for experimental vs control group across whole unsegmented. | 1. There is significant correlation between FO and video segments. 2. There is significant difference in FO of each HN between experimental and control groups. 3. There is significant difference in FO of each HN between experimental and control groups. |
| Viterbi Path | 1. Inter-group differences via **Permutation Testing** and **Hamming distance**: Get mean and SD Hamming distance between experimental and control group participants. Recompute mean and SD Hamming distance for each random permutation out of a total of 5000. Calculate p-value via t test. 2. Intra-group variability: Compute pairwise Hamming distances between all sequences within each group. Find Mean and SD of experimental vs control groups. Do t test/Welch’s t test for significance of difference in Mean. Do F test for significance of difference in Variance. | 1. There is significant difference in Hamming distance between groups compared to random Hamming distance 2. There is significant difference in the mean and variation of Hamming distances between experimental and control groups |
| Switching Rates (SR) | 1. Perform t test between experimental vs control groups. 2. For experimental group, calculate Pearson’s correlation coefficient between SR and PANSS, HDRS, CGI-S AND SOFAS | 1. There is a significant difference between mean and switching rates between the two groups. 2. There is a significant correlation between switching rates and higher scores. |
| Heart Rate  Can also look at amplitude (max and min, arbitrary number, how much blood is rushing in the finger) – difference, can measure change based on baseline  Pulse oximeter outputs 0 – 1 volts (1 max intensity, 0 nothing); box converts that into a 12 bit number | 1. Find mean and variance of HR for experimental vs control groups, divided by:  * Segments * Whole video   Perform t test and F test.   1. Find mean and variance of HR for pooled data, divided by:  * Segments * Whole video   Perform t test and F test.   1. PPG amplitudes?   18. Significant change in HR between segments, or at start compared to end of video? Is there change significantly different between experimental vs control? | 15. There is a significant difference in HR between mean and experimental groups  16. There is a significant difference in HR between segments |

## References

CHANG, L. J., YARKONI, T., KHAW, M. W. & SANFEY, A. G. 2013. Decoding the role of the insula in human cognition: functional parcellation and large-scale reverse inference. *Cereb Cortex,* 23**,** 739-49.

MEER, J. N. V. D., BREAKSPEAR, M., CHANG, L. J., SONKUSARE, S. & COCCHI, L. 2020. Movie viewing elicits rich and reliable brain state dynamics. *Nature communications,* 11**,** 5004.