# 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” analysis project. 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 while they viewed the video stimulus. 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

## Pre-processing

### Cleaning

Scans were normalised to 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

### Determining HMM model parameters

We will fit an HMM Model (HMM-MAR package) on a hold-out dataset (HOD) used for analysis. For this purpose 10 clinical and 10 control participants from the main dataset will be randomly assigned to the HOD. The parameter settings specified in Table 1 will be used for this first model fit.

**Table 1HMM 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 combination of parameter values we will compute how well it predicts the HOD using the following model fit metrics:   * Predictive log-likelihood; use learned parameters to evaluate likelihood of the hold-out sequence * Free energy; Evaluate the variational free energy on unseen data   We will then select the DirichletDiag that maximises the predictive log-likelihood based on the HOD. |
| 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 assumptions are fulfilled (e.g. normalcy, homoscedasticity). If unfulfilled, an updated analysis plan will be uploaded and changes will be reported in Table 2 of the update.

Analyses will be conducted using Network-Based Statistics Toolbox (<https://www.nitrc.org/projects/nbs/>), as well as MATLAB, and R toolboxes / functions.

**Table 2 Outputs of interest for analysis from HMM Model and analysis to be performed**

|  |  |
| --- | --- |
| Output | Hypotheses and Analyses |
| Hidden States (HS) | **H1: Higher PANSS, HDRS, CGI-S and SOFAS scores are significantly correlated with increased FO in ambiguous states.**   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 clinical 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. |
| Average state paths | **H2: There is a significant difference in consistency between groups.**   1. Calculating average state paths using sliding window of 3 consecutive BOLD volumes for each segment and identify the most frequently expressed state (“main HS”) as well as the number of participants that expressed each Hidden State at least once. 2. Calculate **consistency** as % of participants expressing the main HS overall. 3. Analyse intergroup differences of main HS consistency using a Mann-Whitney U Test (or Wilcoxon Rank-Sum Test). |
| Transition Probability | **H3: There are significant differences in transition probabilities between experimental and control groups.**   1. Apply a threshold of 20% to identify the most frequent transitions, and visualise this. 2. Analyse differences between the clinical and control group for each transition probability (# HNs x # HNs matrix) using t-tests. |
| Fractional Occupancy (FO) | **H4: There is significant correlation between FO and video segments.**  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.  **H5: There is significant difference in FO of each HN between clinical and control groups.**  Take the segmented data and calculate FO of each HN for each segment within each group. Conduct **MANOVA**. If a significant effect is foundinterpret post-hoc t-tests*.*  **H6: There is significant difference in FO of each HN between experimental and control groups.**  Do the same as 9. for clinical vs control group across whole unsegmented. |
| Viterbi Path | **H7: There is significant difference in Hamming distance between groups**  Get mean and SD Hamming distance between clinical and control group participants. Recompute mean and SD Hamming distance for each random permutation out of a total of 5000. Analyse using t test.  **H7: There is significant difference in the mean and variation of Hamming distances between experimental and control groups**  Intra-group variability: Compute pairwise Hamming distances between all sequences within each group. Find Mean and SD of clinical 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. |
| Switching Rates (SR) | **H8: There is a significant difference between mean and switching rates between the two groups.**  Perform t-test on SRs between clinical vs control groups.  **H9: There is a significant correlation between switching rates and higher scores.**  For clinical group, calculate Pearson’s correlation coefficient between SR and PANSS, HDRS, CGI-S AND SOFAS |
| Cardiac measures  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 | **H10: There is a significant difference in cardiac response between mean and experimental groups**  Extract HR and PPG amplitudes for each segment and analyses difference between groups using t-tests.  **H10: There is a significant association between cardiac measures and the different segments in the video.**  18. Significant change in HR between segments, or at start compared to end of video? Is there change significantly different between experimental vs control? |

## 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.