Dynamic Bayesian Network for Emotion Modeling

Sarah Brown[1,2], S.R. Prakash[1], Andrea Webb[1], Jennifer Dy[2]

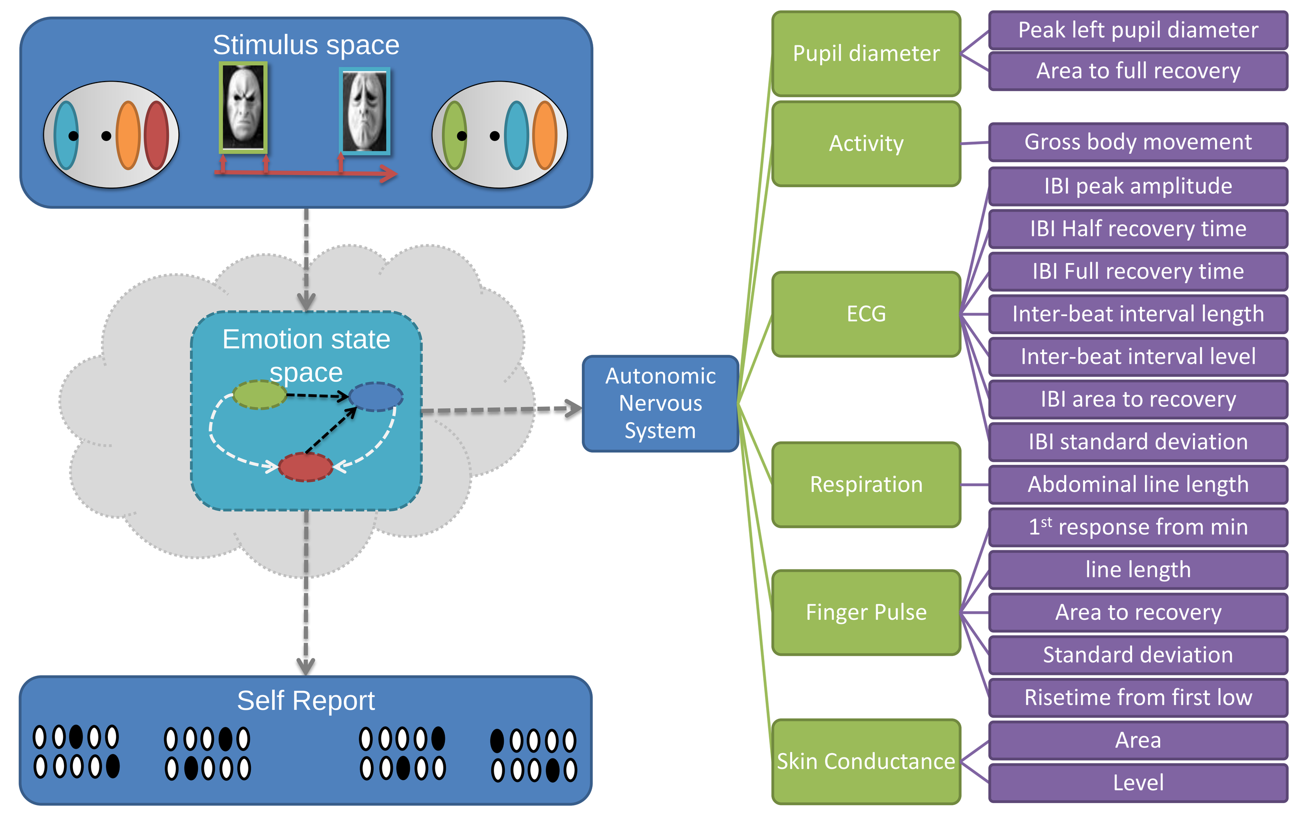
[1] Fusion, Exploitation, and Inference Technologies Group - The Charles Stark Draper Laboratory, [2]Electrical and Computer Engineering – Northeastern University

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-0946746.

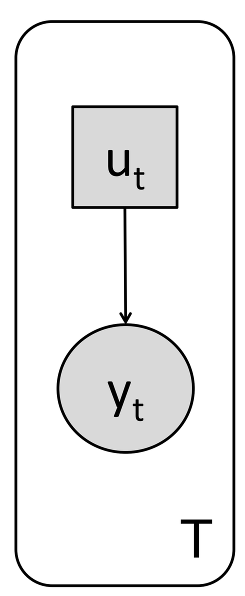
Introduction

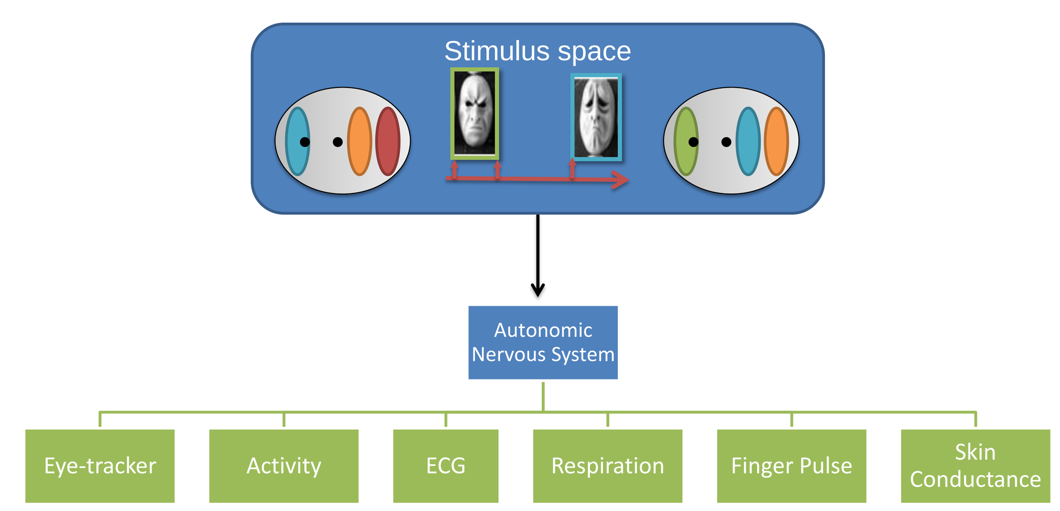
Abstract:

Psychology research frequently employs a variety of statistical methods to analyze data and draw conclusions based on experimental results. The same classifiers, regression models, and clustering techniques that form the basis of study in machine learning are abundant in psychology literature. These statistical methods are typically applied as a black box in their analysis. Through the collaborative efforts of machine learning and psychology researchers, we aim to push the frontiers of psychology research by building on the large and constantly growing body of more powerful machine learning algorithms and by appropriately designing machine learning models that capture modern beliefs about the underlying psychological processes in question. Based on proper modeling and these advances, the power of machine learning to quantify and mechanize qualitative assessments can enable an even greater clinical impact. We explore the implications of machine learning on psychology research specifically in modeling emotion through measurements of peripheral physiology. Of particular interest is the ability to study emotion as a dynamic process. Traditionally psychology data has been analyzed under the assumption of statistically independent trials. This work explores the application of a variation of the hidden Markov model to a dataset collected under a common psychological paradigm. We demonstrate that temporal dependencies in the data are important through improved prediction accuracy of experimental variables. This methodology allows for incorporation of more of the theory about the underlying processes into the data analysis.



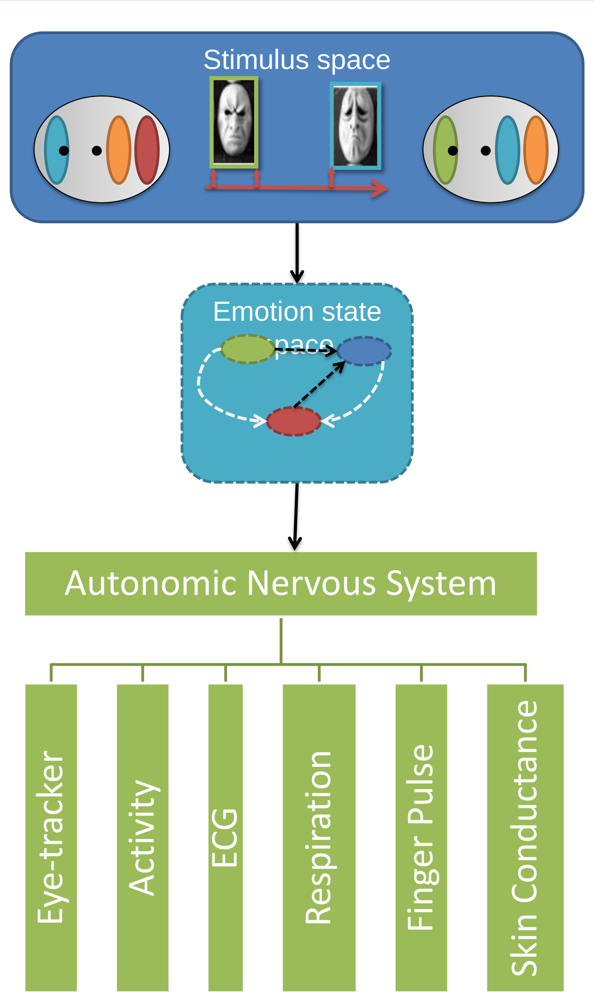
Experimental setup model





Graphical and experimental models implied by classifier based methods commonly employed

Experiment & Model



32 participants (ages 19-55, 53% male and 47% female) were each presented with two sequences, one of sounds (Bradley et al, 2007) and one of images (Lang et al, 1999), each consisting of 24 emotionally evocative stimuli. The same sequence of stimuli was used for each subject. After each stimulus presentation, the subject indicated the degree to which they experienced each of 5 discrete emotions (fear, disgust, amusement, sadness, and anger) on a scale of 1 to 7. Stimuli were chosen to represent these 5 discrete emotions as well as neutral stimuli from standard databases used for emotion research. During the stimulus presentations, the following physiological signals were continuously recorded: Electrocardiogram (ECG), Pupil Diameter (via eye-tracker), Skin Conductance, Respiration, Finger Pulse, and Activity (gross body movement)

Features were extracted from the physiological waveforms using CPSLAB (Scientic Assessment Technologies, Salt Lake City, UT). Though 23 features were extracted from the 6 raw signals, feature selection was performed to choose optimal subsets of 3 features for each sound and image stimuli.

We propose applying an HMM with input, the transition model is multinomial and we apply a Gaussian output model. The proposed graphical and implied experimental model are shown. The key differences rom traditional analysis are a time dependency and introduction of an abstract latent state that relates the stimulus and response.

Joint Probability Model:

$$P(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{U}) = \prod_{s=1}^{S} \prod_{t=1}^{T} \prod_{i=1}^{K} \left(P(\boldsymbol{y}_{s,t} \mid x_t = i) \prod_{j=1}^{K} \left(\prod_{k=1}^{K} t_{i,j,k}^{[x_{s,t-1}=k]} \right)^{[u_{s,t}=j]} \right)^{[x_t=i]}$$

Transition matrix

$T = \{t_{i,j,k}\}$ $\quad t_{i,j,k} = P(x_{t+1} = k \mid x_t = j, u_{t+1} = i)$

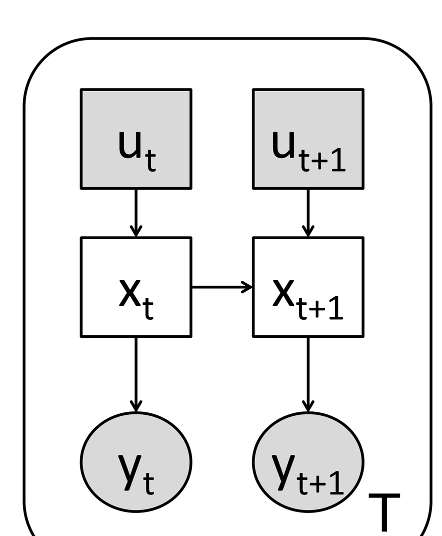
$i \in \{amusement, disgust, sad, anger, fear, neutral\}$

$\text{j, k} \in \{ \textit{discrete abstract mental states,} \}$

Physiological State Descriptions

$P(y \mid x = k) = \mathcal{N}(\mu_k, \Sigma_k)$

$\mu_k \in R^D, D\ measurements$ $\quad \Sigma_k \in R^{D \times D}$

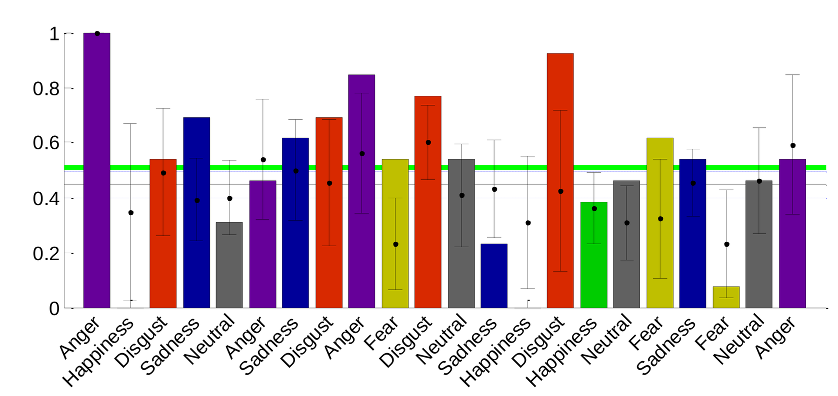


Results

All accuracy results are based on subject-wise cross validation of a stimulus class prediction task. After training, the IOHMM was tested by predicting the stimulus label. Accuracy rates are to the right. It is compared with two standard classifiers. The 5-class problem for the classifiers neglects neutral stimuli as is common in analysis, however due to the temporal nature the IOHMM cannot be trained completely without so no 5 class problem is presented. The proposed model out performs the classifier and as shown below provides additional insights to the problem.

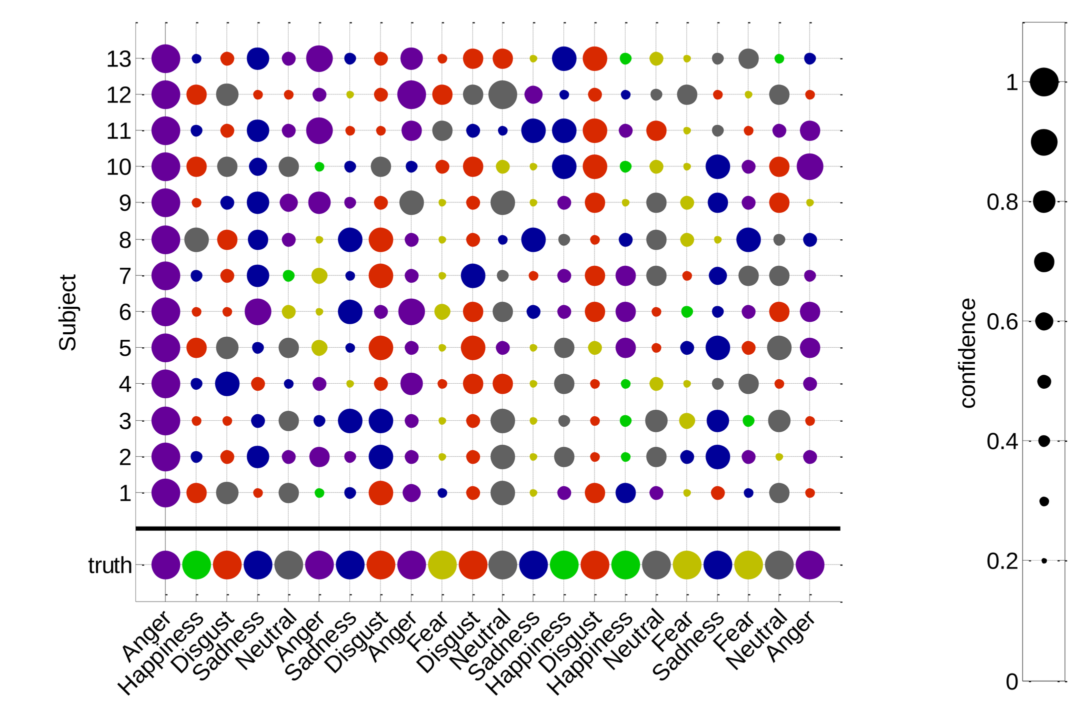
Model & cross validation	Sound Stimuli	Image Stimuli
LDA 5 class 50/50MC	0.3894	0.3133
SVM 5 Class 50/50MC	0.3022	0.2880
LDA 6 Class 50/50MC	0.3167	0.2815
SVM 6 Class 50/50MC	0.2086	0.2333
IOHMM 50/50MC	0.4069	0.4462
IOHMM LOO	0.537	0.5629

Sound Stimulus Prediction Accuracy



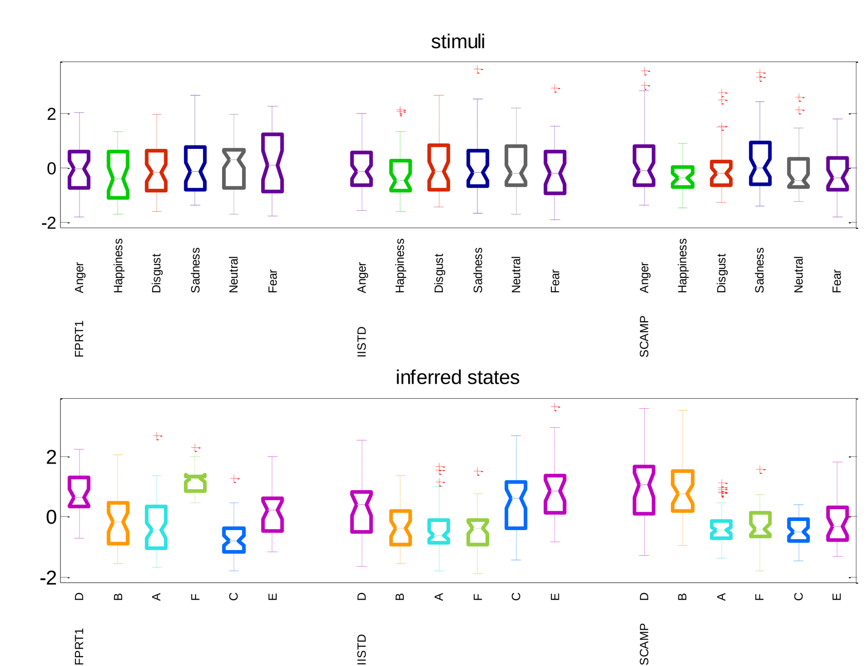
This figure shows accuracy per stimulus in the sequence, stimuli of the same label are shown in the same color and the average accuracy across stimuli is indicated by the horizontal green line behind the bars. The error bars represent the variation in the accuracy of of predicting each stimulus across various random train-test partitions. The horizontal black line shows the average accuracy across all cross validation trials as shown in Table\ref{tab:pAccStim} and the blue dashed lines are the standard deviation in total accuracy. These results are for the trial with the highest total accuracy with image stimuli.

Sound Stimuli Prediction Confidences by Subject



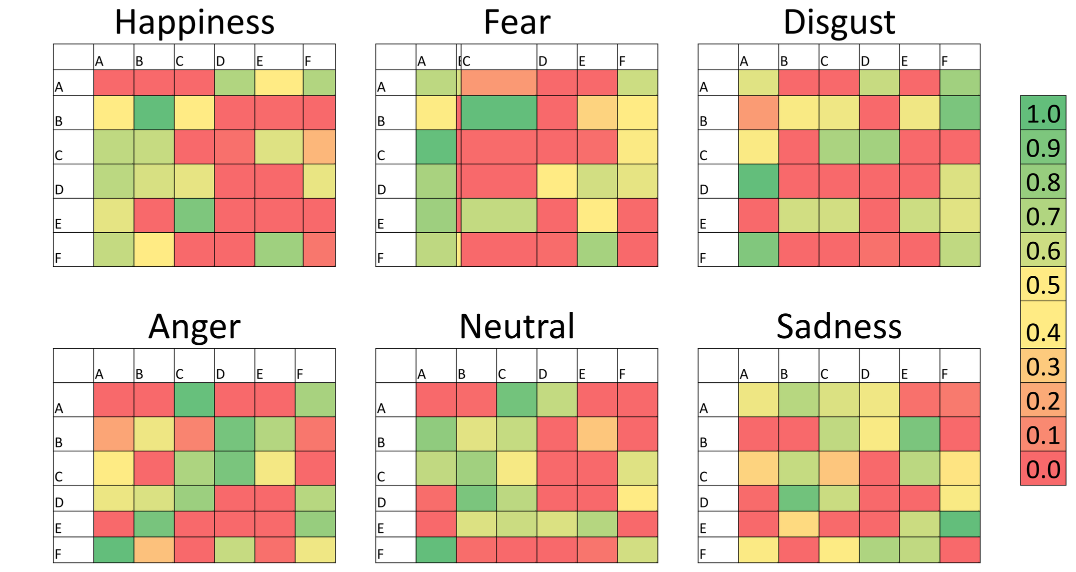
This shows the confidence in the stimulus class prediction for the best performing repetition in a 50-50\% Monte Carlo cross validation. The truth row shows the true class of each stimulus. The color of the circle indicates predicted stimulus class and the diameter indicates the confidence. The right shows the scales as a reference for the confidence levels

Sound Stimuli Physiology of Test Subjects



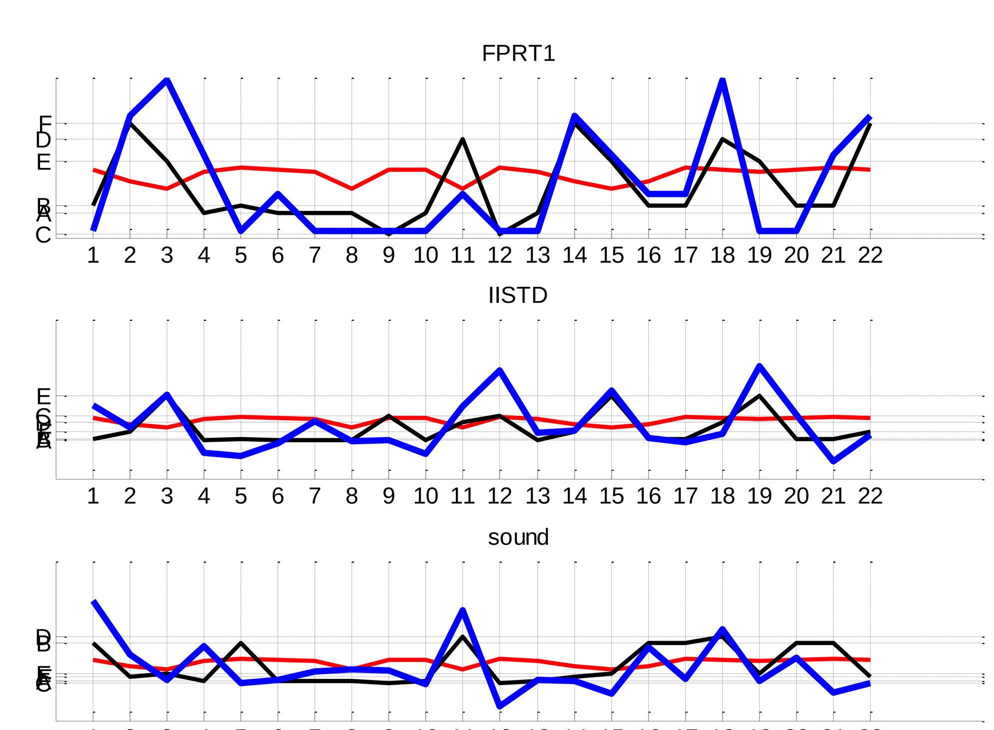
Distribution of physiology per feature by \(top\) stimulus class and \(bottom\) learned, abstract state. The horizontal line in each box is the median. The box limits represent the 25^{th} and 75^{th} percentiles and whiskers indicate the full range of points not considered to be outliers. Outliers are noted with +'s. Two medians are significantly different at the 5\% level if the notch intervals do not overlap. These data are from the test subjects for the highest accuracy cross validation trial.

Sound Stimuli Transition Matrices

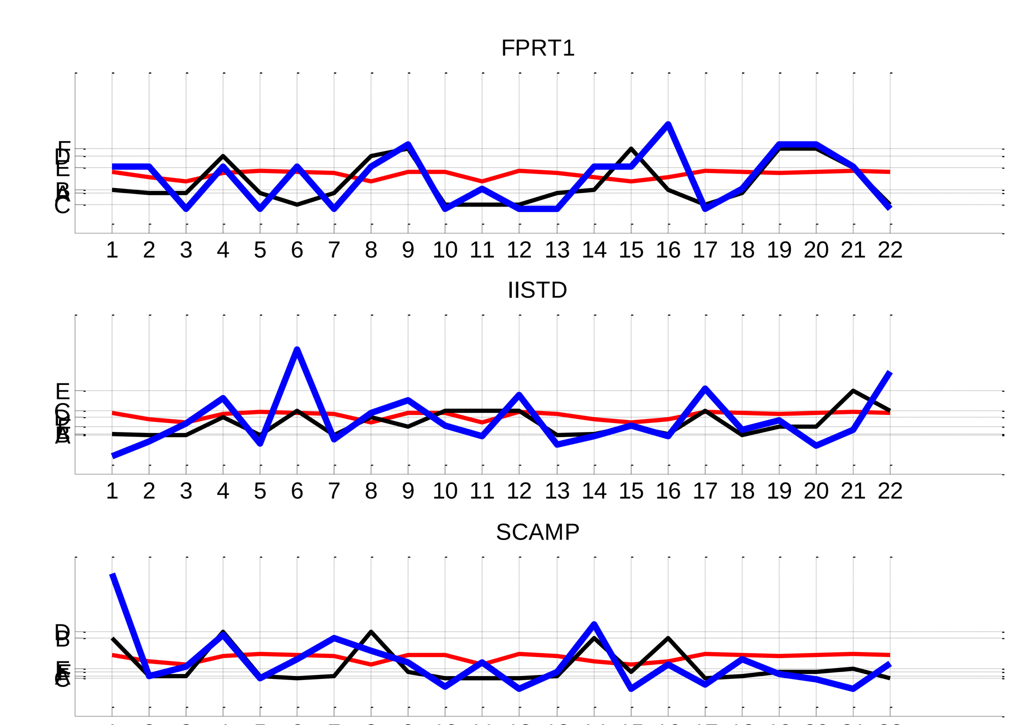


Heatmaps of the transition tables for each stimulus time for the best performing 50\%-50\% MC cross validation trial. Each cell represents the probability of transitioning from the state given by its row, to the state given by its column, when shown a stimulus labeled as given by the table.

Most Accurate (68.18) Subject Physiology



Least Accurate (40.91) Subject Physiology



Sample true (blue) expected based on abstract states (black) and expected based on stimuli alone (red) physiology for two subjects. The IOHMM is a better generative model and more capable of explaining the observed physiology as demonstrated through the black traces being a closer estimate to the true physiology than the red estimates. This also demonstrates that based on strictly stimulus labels the physiology classes are not very different..

Northeastern

DRAPER LABORATORY