

Characterization of brain states during task-switching using a neural network classifier. G30

A. Lenartowicz, G. Detre, S. Polyn, J. Chein, N. Yeung*, L. Nystrom, K.A. Norman, J.D. Cohen
Princeton University, Princeton, NJ, *Carnegie-Mellon University, Pittsburgh, PA

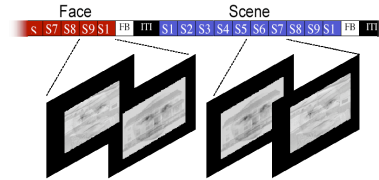
INTRODUCTION

The pattern of neural activity present at a given time may be seen as a reflection of the cognitive and behavioral state of the individual. In the present work, we trained a neural network algorithm to classify patterns of fMRI data from a task-switching experiment on a TR by TR time scale (approximately 2 sec.) (Hanson et al., 2004; Polyn et al., 2004). We explored activity in the entire cortex, as well as activity constrained to the prefrontal cortex, as well as activity constrained to the prefrontal cortex. The latter analysis was motivated by biased competition theories (Cohen & Miller, 2001) which posit a key role for PFC task representations in cognitive control. The classifier succeeded at matching up neural states to the correct task state (accuracy >.85), even when tested on single TRs; this holds for the entire cortex and the PFC. Preliminary analyses relating classifier performance to behavior are discussed.

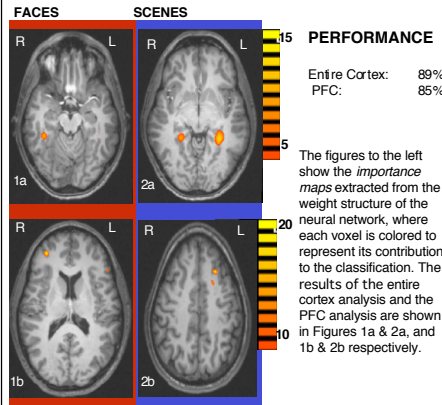
METHODS

- Experimental data are described under Dataset 1 and Dataset 2. Both studies involved task-switching: alternating between face & scene judgments (for stimuli comprised of overlapping faces & scenes) OR face & word judgments (for stimuli comprised of overlapping faces and words).
- Imaging data were fully preprocessed (motion corrected, smoothed, co-registered, normalized to a Talairach template).
- Preprocessed data were z-scored and voxels whose activity was not discriminative between conditions were eliminated from the analysis.
- To classify the data, single-TR brain volumes were fed to a 2-layer network (no hidden layer, trained with backpropagation) that mapped patterns of voxel activity onto two output units, corresponding to the two tasks.
- Leave-one-out cross-validation was used at test: one run was withheld from training and presented (TR by TR) at test. This process was repeated for each run in the session. Classifier performance represents the percentage of TRs presented at test for which the correct output unit was more active.

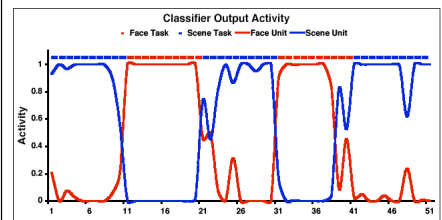
DATASET 1



Brain volumes were collected while participants alternated between deciding if a face is male or female, or if a scene is inside or outside. Only task TRs (face/scene blocks) were included in the classification training set. One run of the experiment was withheld at each training session to be used as the testing set. This dataset (of 1 participant only) was used as pilot data: Could the classifier correctly identify the behavioral task based on neural activity? Which regions would be key contributors to the classification at output?

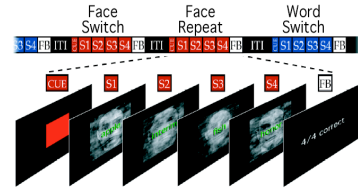


1a: BA19/37 [35 -39 -13] 1b: GH [-27 -40 -3]
2a: BA46 [30 46 15] 2b: BA8 [-23 22 47]
(Talairach coordinates)

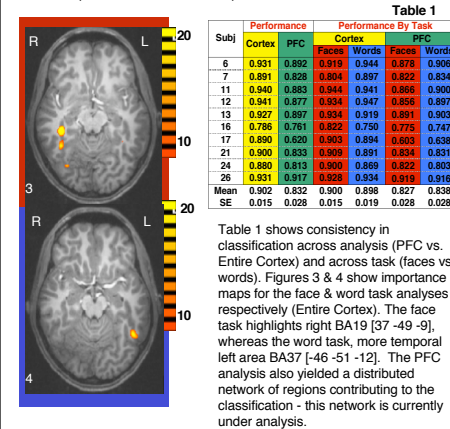


The above time course shows the first 50 TRs (5 blocks) of the task. The broken lines indicate the correct task, the continuous lines show the actual activity in the two output units. Note that that overall the 'correct' unit is generally considerably more active than the 'incorrect' unit.

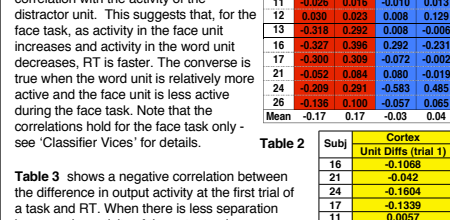
DATASET 2



Brain volumes were collected while participants alternated between deciding if a face is male or female, or deciding if a word is abstract or concrete. The second dataset was selected to: 1. extend our analysis across multiple participants (10 were analyzed) and 2. attempt to correlate output unit activity with behavioral performance (RTs). Pertaining to goal 2: this data was selected because of variability in behavioral performance that was not present in Dataset 1.



3a: BA19/37 [35 -39 -13] 3b: GH [-27 -40 -3]
4a: BA46 [30 46 15] 4b: BA8 [-23 22 47]
(Talairach coordinates)



The above time course shows the first 50 TRs (5 blocks) of the task. The broken lines indicate the correct task, the continuous lines show the actual activity in the two output units. Note that that overall the 'correct' unit is generally considerably more active than the 'incorrect' unit.

CONCLUSIONS

Classifier Virtues:

- We can identify brain activity patterns relating to particular cognitive states on a TR by TR basis
- We can use classifier output as a parametric measure of task state engagement - the correlational trends we identified suggest that we may be able to relate classifier output to behavioral performance
- The combination of TR by TR resolution and correlational approaches may allow us to establish brain-behavior correlations across TRs rather than based on averaging data over the entire session (i.e., correlating across subjects)

Classifier Vices:

- In Dataset 2, we found correlations between RT and classifier output for the face task only (Table 2), and yet classifier performance on the word task is equivalent to that of the face task (Table 1). The network may be coming to a correct classification without using all of the input information. We suggest that this may occur because backpropagation is opportunistic: If the neural response to one task is stronger than the neural response for the other task, then the algorithm may classify based on the absence/presence of the stronger task rather than factoring in both representations. We may be able to get around this problem by using a different classification algorithm (e.g., correlation-based classification, Haxby et al., 2001)

- Although the correlations described here suggest an interesting relationship between classifier output and behavior, there is still room for the correlations to be improved. It remains to be seen which behavioral metric will correlate best with classifier output (e.g., raw RT, differences between various trial types), and whether advanced corrections for the hemodynamic response function will be necessary to optimize the RT to classifier output relationship.

REFERENCES

Hanson,S.J., Matsuka,T., and Haxby, J.V. Combinatoric Codes in Ventral Medial Temporal Lobes for Objects: Is There a Face Area?. *Neuroimage*, 2004.
Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425-2430
Polyn, S. M., Nystrom, L. E., Norman, K. A., Haxby, J. V., Gobbini, M. I., & Cohen, J. D. (2004). Using neural network algorithms to investigate distributed patterns of brain activity in fMRI. Human Brain Mapping conference, Budapest, Hungary.
Polyn S.M., Cohen J.D., Norman K.A. (2004) Detecting distributed patterns in an fMRI study of free recall. *Society for Neuroscience conference*, San Diego, CA.

Acknowledgments:
National Institutes of Health (NIMH)
Grant: 5 R01 MH052864