Deep learning classifiers of visual cortex activity can identify which moment of a video is represented by a single fMRI volume during naturalistic movie viewing

Matthew R. Johnson[a], Jacob M. Williams[a], Rafay A. Khan[b], & Karl Kuntzelman[a]

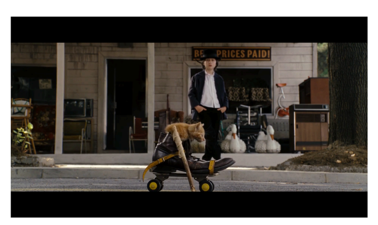
[a]University of Nebraska-Lincoln [b]University of Illinois at Urbana-Champaign

INTRODUCTION

- How accurately can we hope to decode the continuously varying contents of complex, naturalistic thoughts and percepts using fMRI measures of brain activity?
- Most contemporary classification and decoding techniques are limited in flexibility and inferential power.
- Here, we apply a new technique developed by our group called Paired Trial Classification (PTC)[1], which uses deep neural networks to determine whether two brain activity patterns are similar or different.
- We compare PTC to two traditional neural similarity measures for visual cortex activity during naturalistic movie viewing.

TASK DESIGN

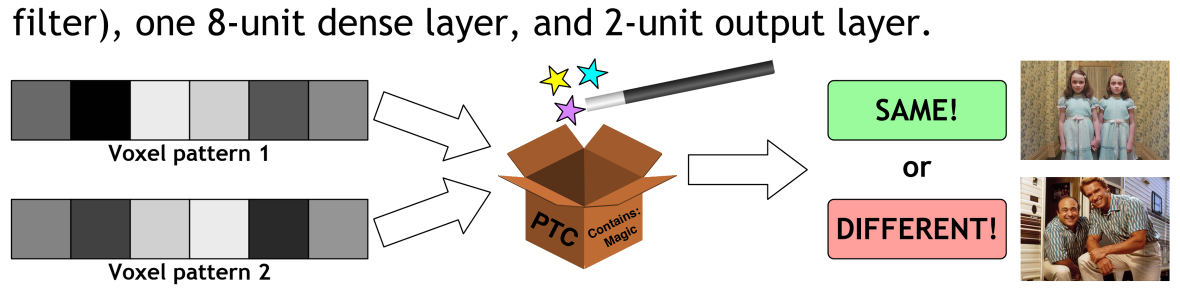
- Participants (N=15) viewed (and listened to) one of two six-minute films, the beginning of the first episode of *Pushing Daisies* or the opening to *The Brothers Bloom*.



- Each viewing comprised one fMRI run. Participants saw the film three times in the scanner, several minutes apart. (With other runs in-between that are not analyzed here.)
- Instruction was simply to watch and listen attentively, and try to remember as much detail as possible.
- 5min runs of resting-state data were also collected at the beginning and end of the scan session.

ANALYSIS METHODS

- Whole-brain multiband fMRI, 2.5mm isotropic voxels, TR=1000ms.
- Pre-processing: Motion correction and linear registration to MNI template. Template then warped into individual-subject space to create ROI masks of visual cortex (via Harvard-Oxford atlas).
- PTC: Deep learning models (using DeLINEATE toolbox[2]) trained on data from the second half of each run. fMRI patterns from two volumes (from different runs) were fed in; PTC classifier was trained to distinguish if the volumes represented the same timepoint in the film, or different timepoints.
- Simple network structure: One convolutional layer (single 2x10 filter), one 8-unit dense layer, and 2-unit output layer.

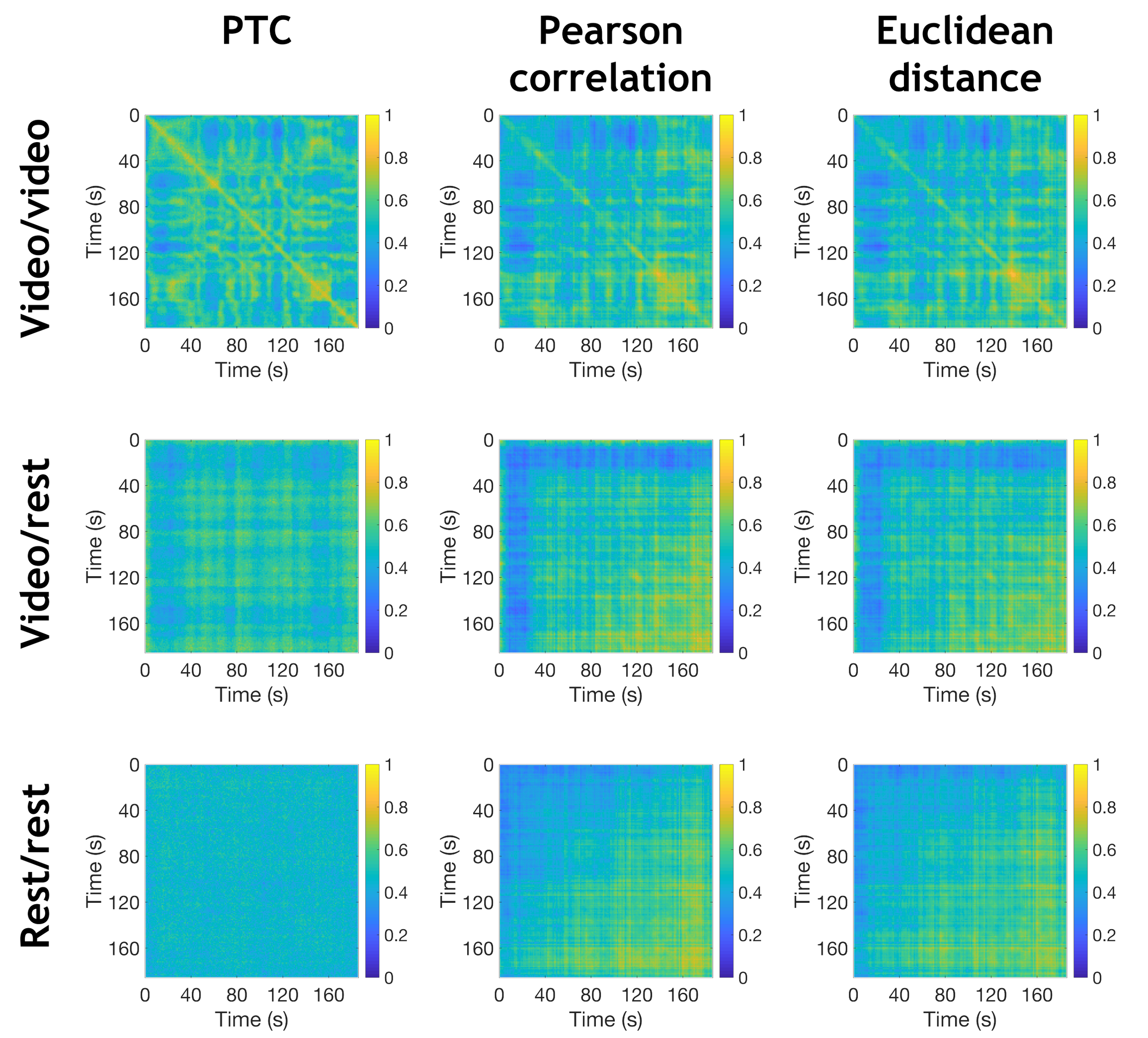


- All similarity measures were then applied to data from first half of each run.
- Traditional pattern similarity measures: Pearson correlation (with Fisher z' transformation) and -1 * Euclidean distance (so for all measures, higher values = more similarity).

RESULTS

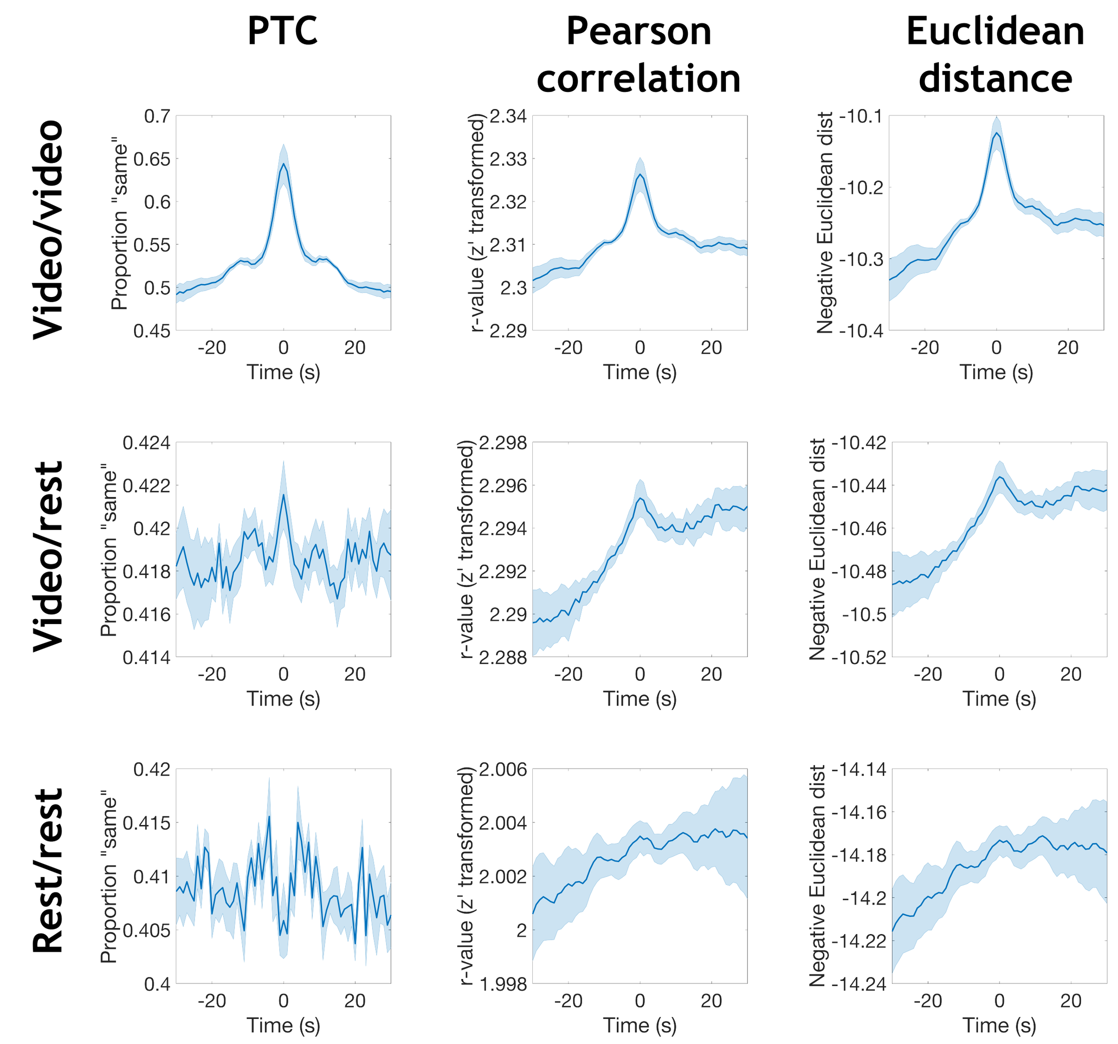
Similarity matrices

- *between each timepoint of one run and each timepoint of another run*
- *averaged over all possible combinations of the different run types (video run/video run; video/rest; rest/rest)*
- *all metrics normalized to 0-1 scale*



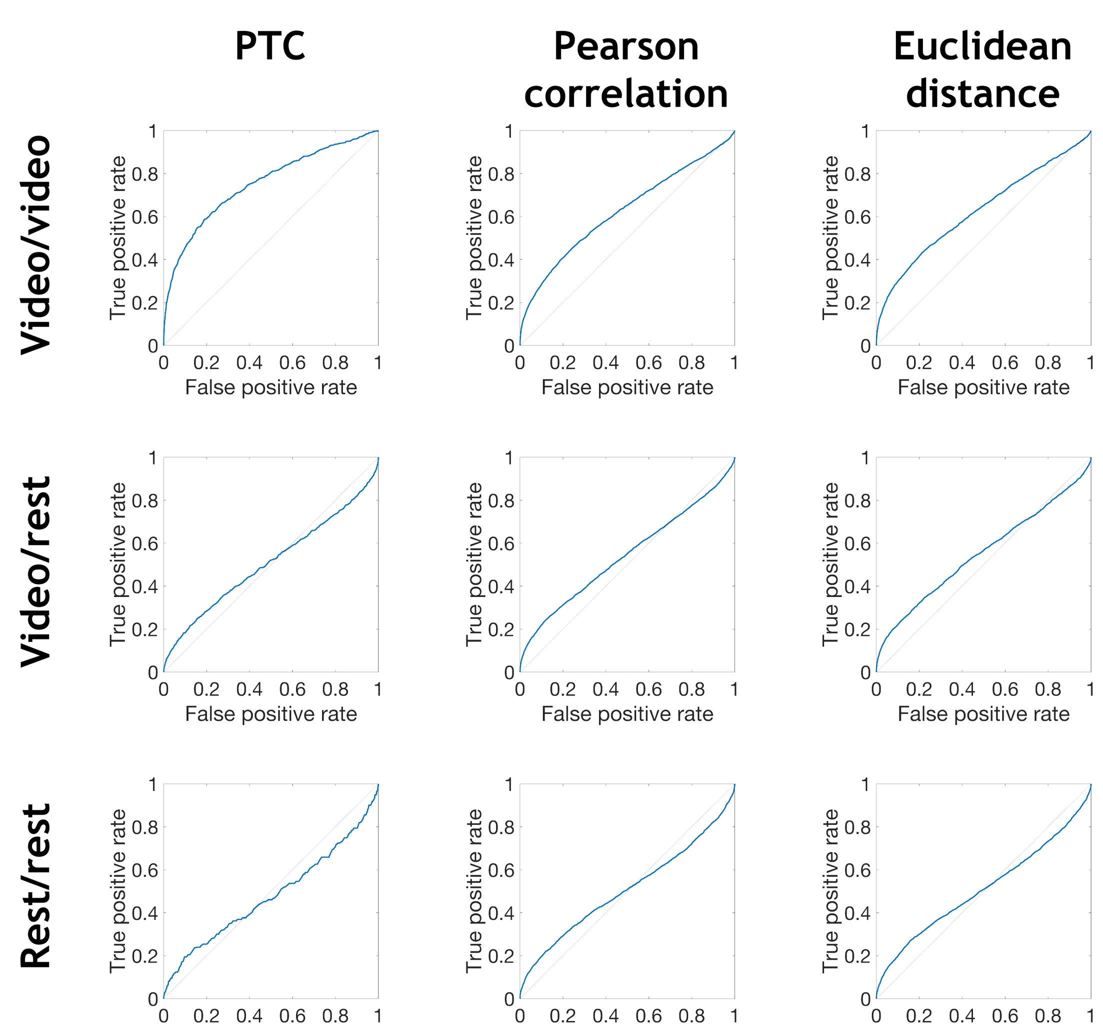
Time-windowed similarity values

- *between each timepoint of one run and all timepoints in a 30-second range of that timepoint in another run*
- *all metrics are shown in the "native" units for that similarity measure*



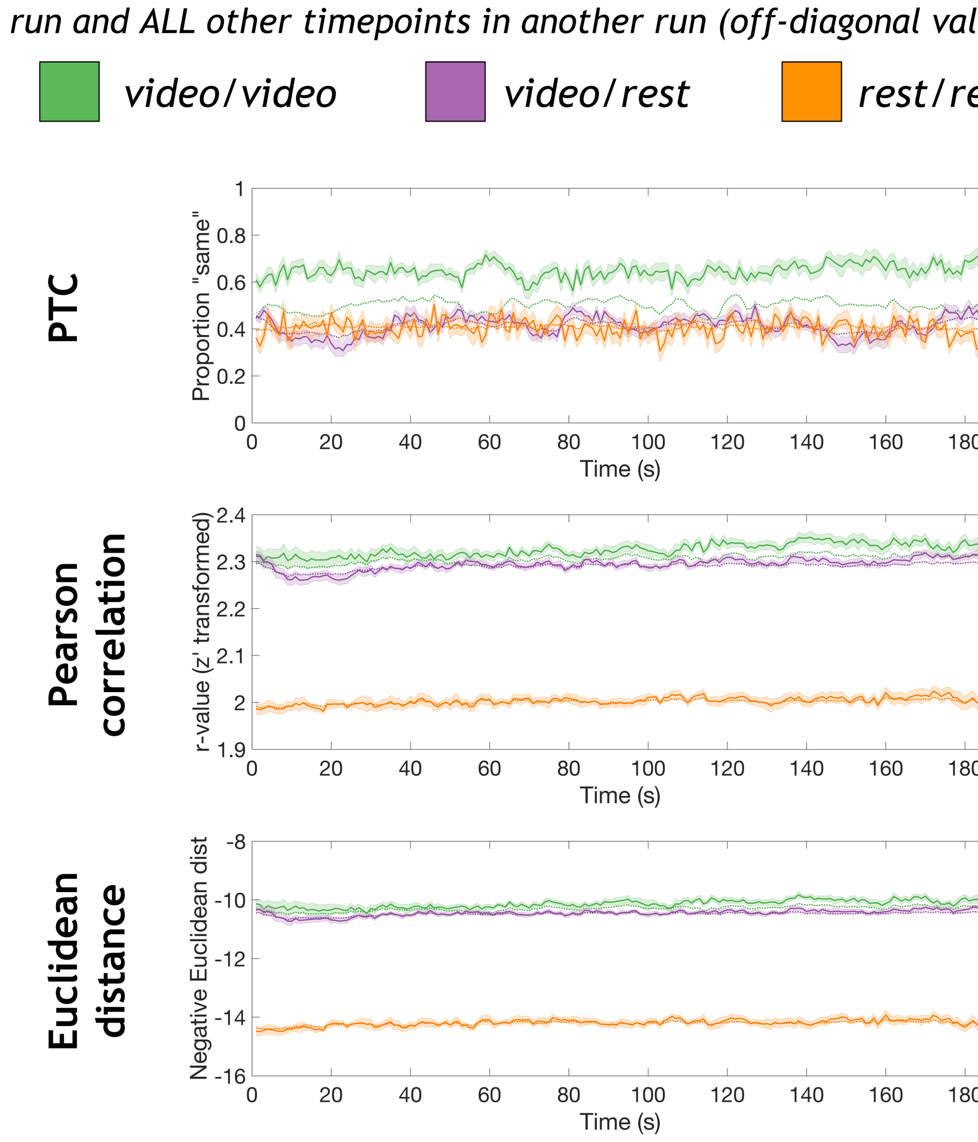
ROC curves

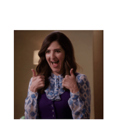
- *generated from similarity scores contained in the matrices above, with a varying discrimination threshold*
- *only on-diagonal value pairs (i.e., exact same timepoint in different runs) were considered "true" positives*



Comparison between run type pairs

- *solid lines: similarity scores between each timepoint in a run and the same timepoint in another run (on-diagonal values)*
- *dotted lines: averaged similarity scores between each timepoint in a run and ALL other timepoints in another run (off-diagonal values)*



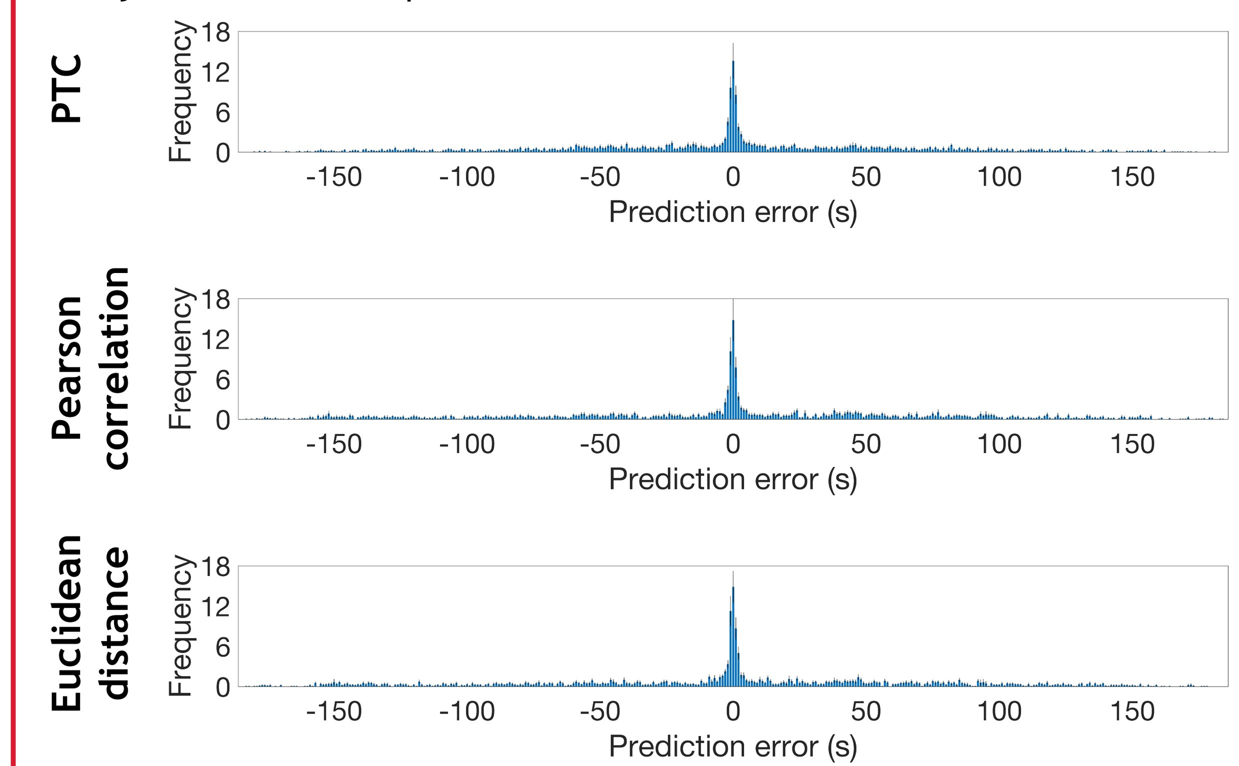


FUN FACT! Overall PTC accuracy during training was 69.7% - which breaks down into 76.5% accurate for "same" voxel pattern pairs (drawn from the same timepoint in different runs) and 62.8% accurate for "different" voxel pattern pairs!

RESULTS (cont.)

"Best guess" analysis for each timepoint

- *use the highest similarity scores in the matrix to generate a "best guess" of each voxel pattern's position in time, based on its similarity to timepoints in the other runs*
- *then plot histograms of how far off all those guesses were*
- *only video/video run pairs are shown here*



CONCLUSIONS / FUTURE DIRECTIONS

- All measures generated reasonable similarity matrices and had fairly good predictions in the "best guess" analyses.
- However, Pearson correlation and Euclidean distance were more heavily influenced by temporal nuisance factors (e.g. scanner drift) to which PTC was insensitive.
- Thus, PTC overall had more predictive power than either of the more traditional analysis types (cf. ROC curves).
- Other advantages of PTC: Can be used to deliver either continuously varying similarity scores or binary decisions. In both cases, the values stay within a 0-1 range, which is more convenient for human interpretation.
- Future directions: Apply not just to continuous perception, but to a continuous stream of mental imagery data. Spoiler alert: It works on that, too! (If you're going to the Vision Sciences Society meeting in May, we'll be talking about that there.)
- Past directions: If you have a time machine, go back to poster B102 from yesterday and see our presentation on this general technique. Try not to step on any butterflies. (Or just keep an eye out for our manuscript, coming soon. Soon...ish.)

REFERENCES / ACKNOWLEDGEMENTS

[1] Williams JM, Samal A, Rao PK, Johnson MR. 2019. Paired Trial Classification: A novel deep learning technique for MVPA. *Cognitive Neuroscience Society 26th Annual Meeting*, poster B102.

[2] Kuntzelman K, Williams JM, Samal A, Rao PK, Johnson MR. 2019. DeLINEATE: A deep learning toolbox for neuroimaging data analysis. Cognitive Neuroscience Society 26th Annual Meeting, poster B105. http://delineate.it.

Supported by NSF/EPSCoR grant #1632849 to MRJ and colleagues.