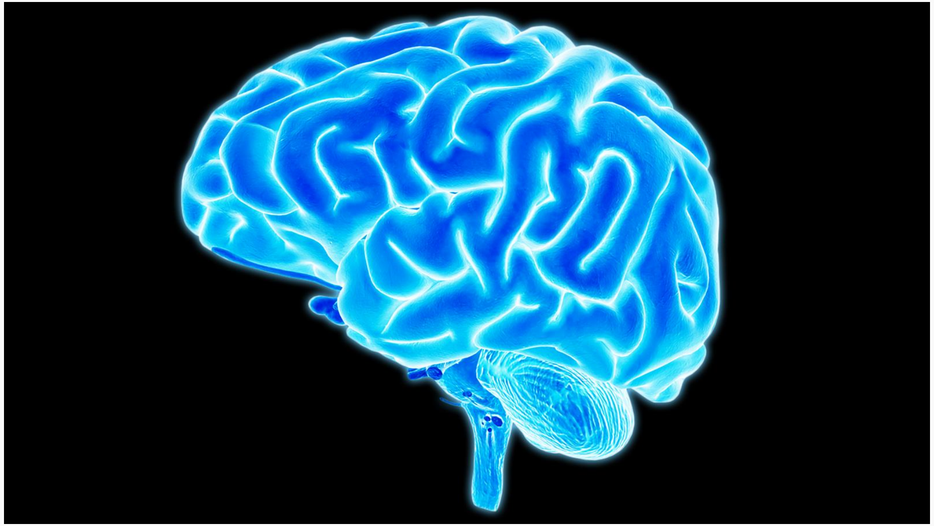
Mapping Between Natural Movie fMRI Responses and Word-Sequence Representations**

Kiran Vodrahalli*, Po-Hsuan Chen*, Yingyu Liang*, Janice Chen*, Esther Yong † , Christopher Honey † , Peter J. Ramadge*, Kenneth A. Norman*, Sanjeev Arora*
NIPS MLINI 2016
December 9, 2016
** work supported by Intel Labs (www.brainiak.org)
* = Princeton, † = U. Toronto

fMRI: Sensing Brain Signal



100 billion neurons in the brain

fMRI measures hemodynamic response at ~10^5 different 3mm x 3mm x 3mm voxels

Each voxel represents an average of the activity of the ~10^6 neurons it contains

Goal: **detect semantic meaning in this signal.**

Prior Work on Connecting a Semantic Space to fMRI Data

[Mitchell et al ‘08] predicts fMRI responses induced by **pictures of concrete nouns**.

[Naselaris et al ‘09] predicts fMRI responses induced by **images of scenes.**

[Pereira et al ‘11] uses the same dataset as Mitchell ‘08, but focuses on **generating words** related to the concrete nouns.

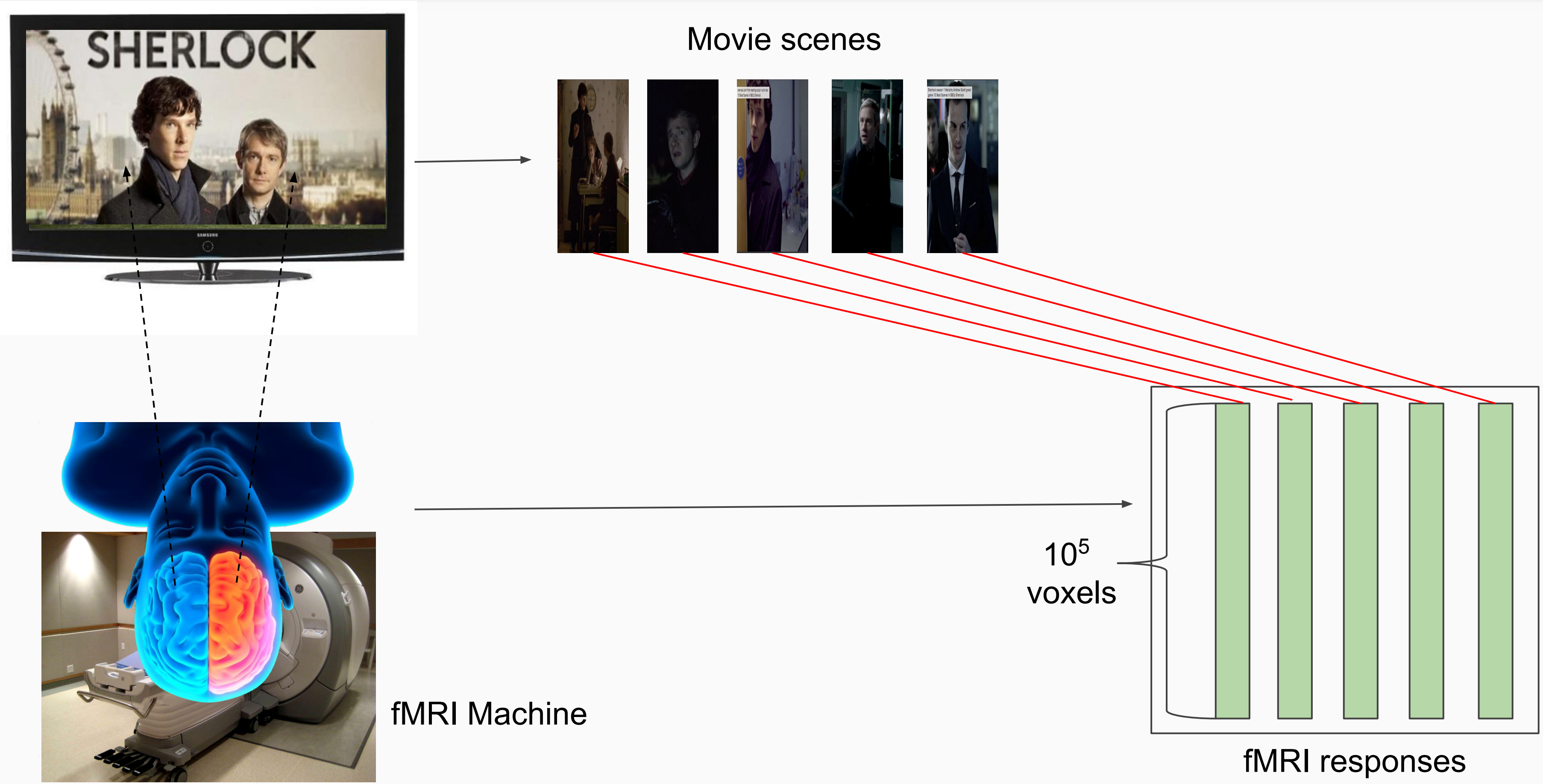
[Naselaris et al ‘11] tries to **reconstruct movie images** from fMRI signals measured while subjects watched movies.

[Wehbe et al ‘14] has subjects **read a chapter of Harry Potter** and predicts fMRI responses for held-out time points.

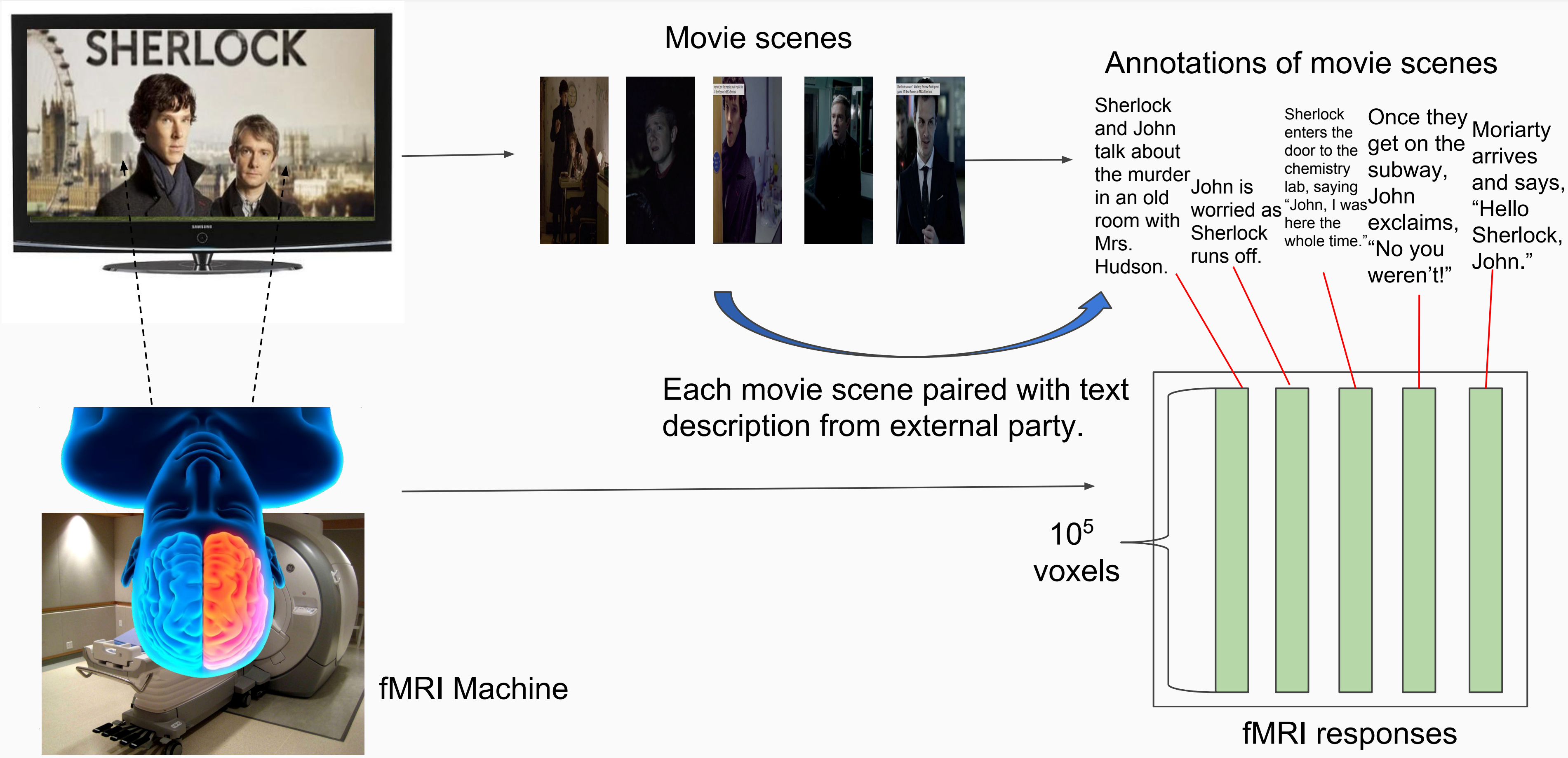
[Huth et al ‘16] reconstructs fMRI responses to **auditory stories**.

[Pereira et al ‘16] decodes fMRI responses to **word clouds and short sentences.**

Goal 1: Decode fMRI Response Semantics



Goal 1: Match fMRI responses to annotations (Views: fMRI signal, text annotations)



Interesting and Useful Discoveries

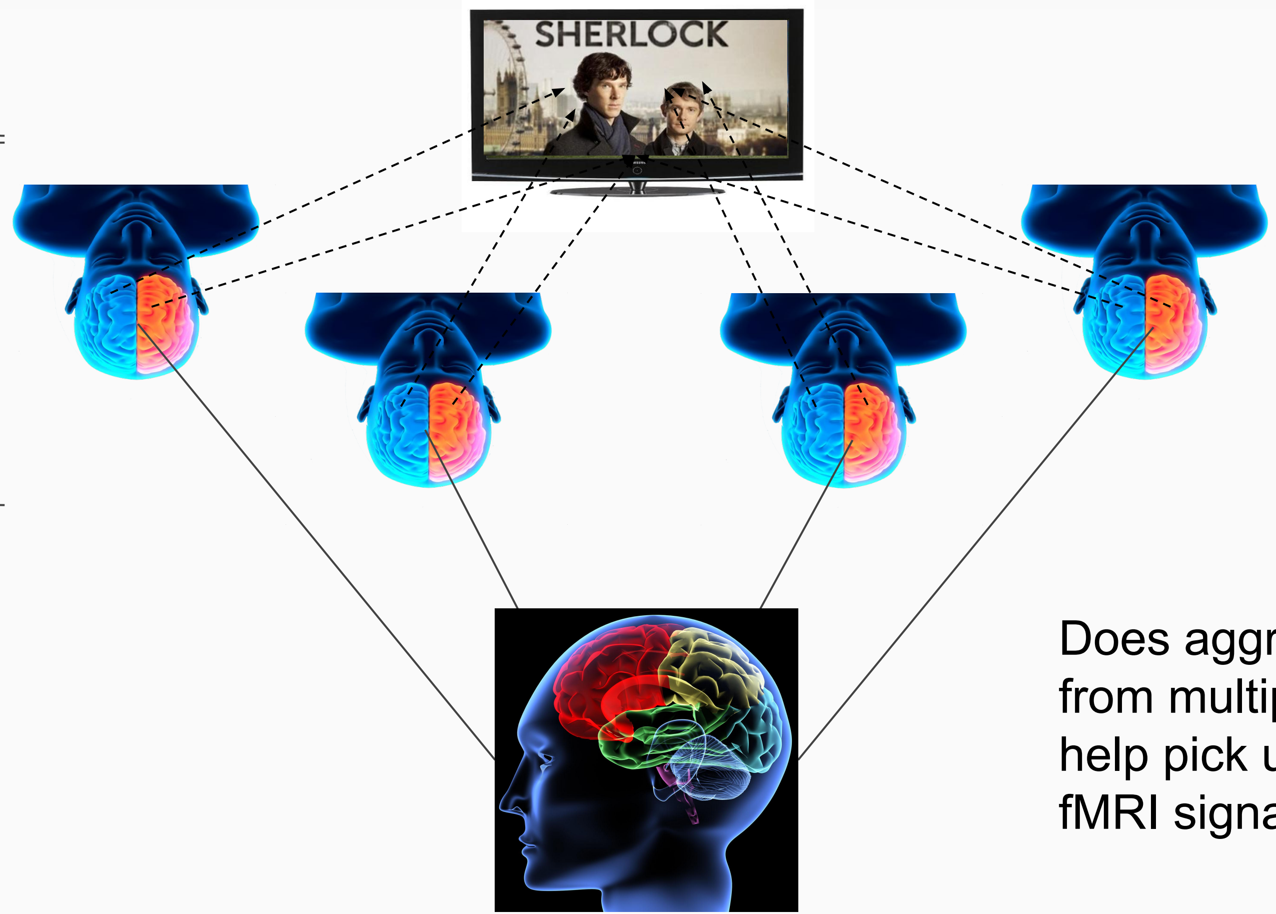
- The Shared Response Model (SRM, Chen et al. 2015) helps for decoding text!
- Weighted average word vectors → better semantic context vectors (ICLR 2017 submission, Arora et al)
- Orthogonal maps decode fMRI → text better than ridge regression

Goal 2: Leverage Multiple Subject Views to Extract Better Semantics

Shared Movie Stimulus

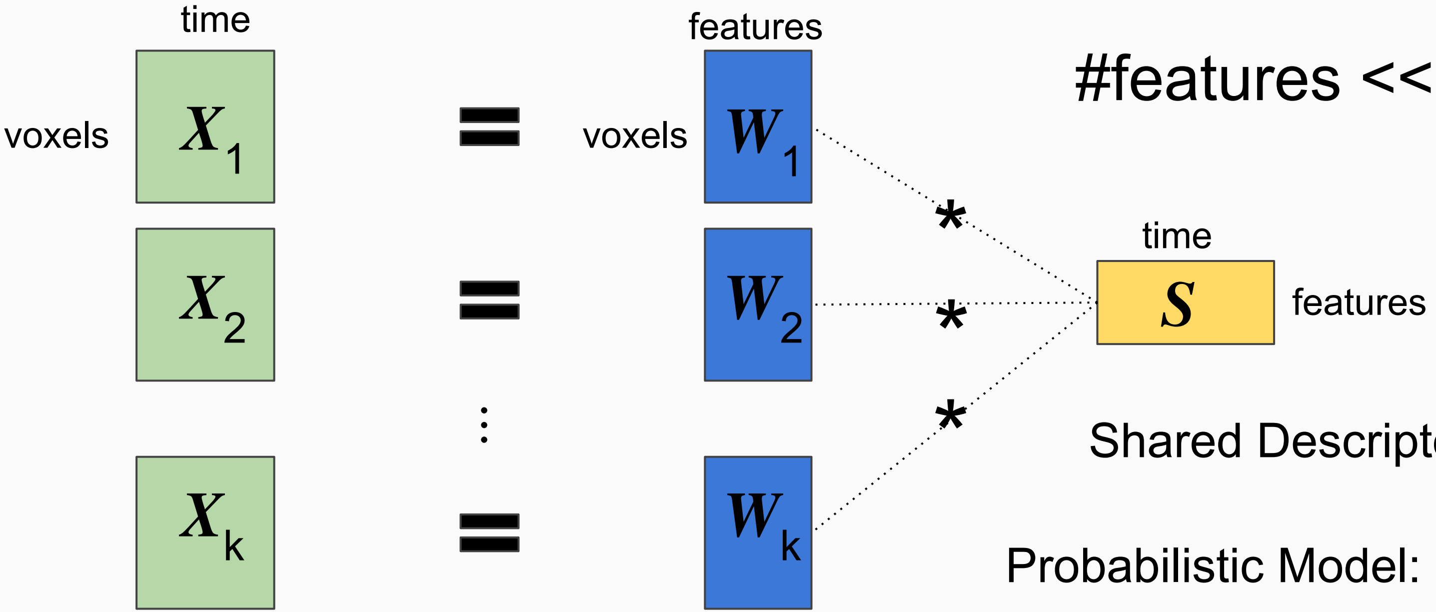
Multiple Subject Responses

Shared fMRI Response



Does aggregating data from multiple individuals help pick up a stronger fMRI signal?

Shared Response Model (SRM, [Chen, Chen, Yeshurun, Hasson, Haxby, Ramadge '15])



#features << #voxels

Shared Descriptor of Semantic Signal

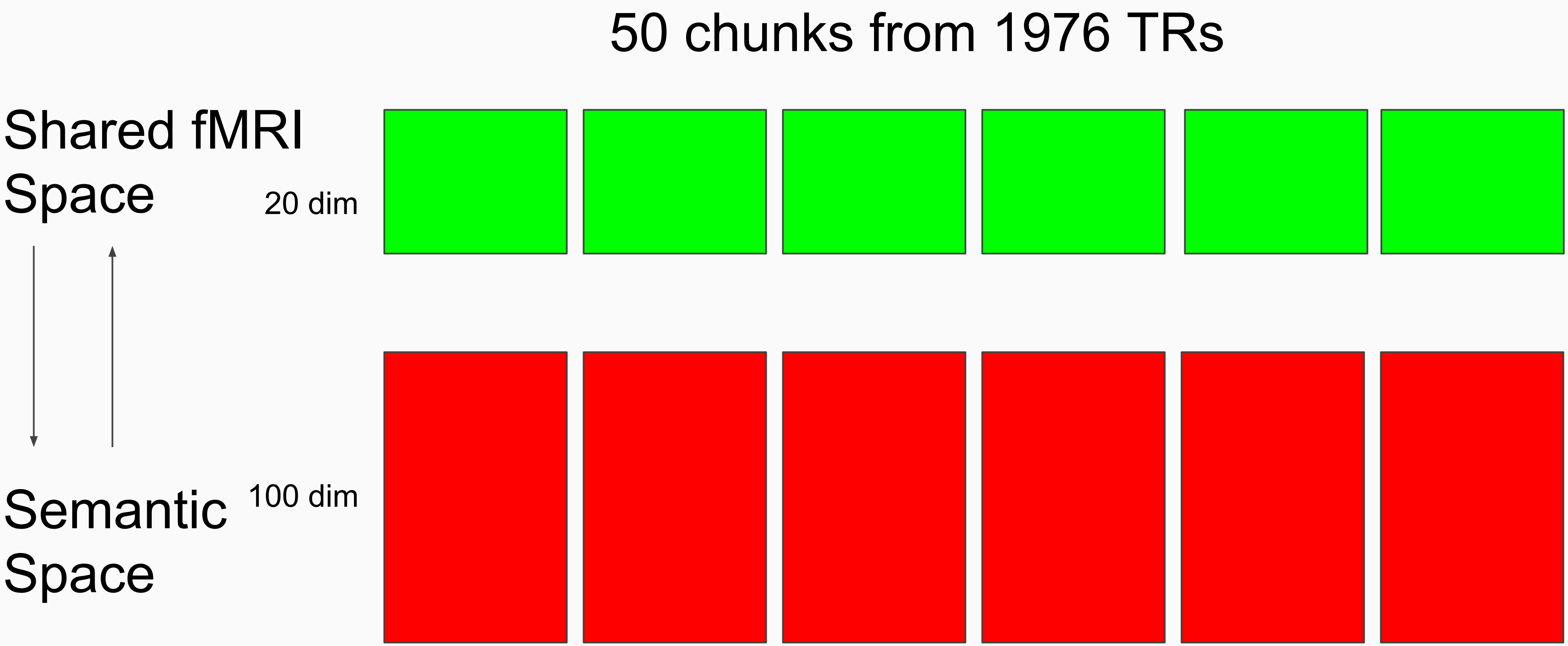
$$\underset{W^TW=I;S}{\operatorname{argmin}} \sum_{i=1}^{k} \|X_i - W_i S\|_F$$

Probabilistic Model:

$$s_t \sim \mathcal{N}(0, \Sigma_s)$$

$$x_{it} | s_t \sim \mathcal{N}(W_i s_t + \mu_i, \rho_i^2 I)$$

Scene Classification/Ranking Experiments



Results - Top-20% Classification and Average Rank

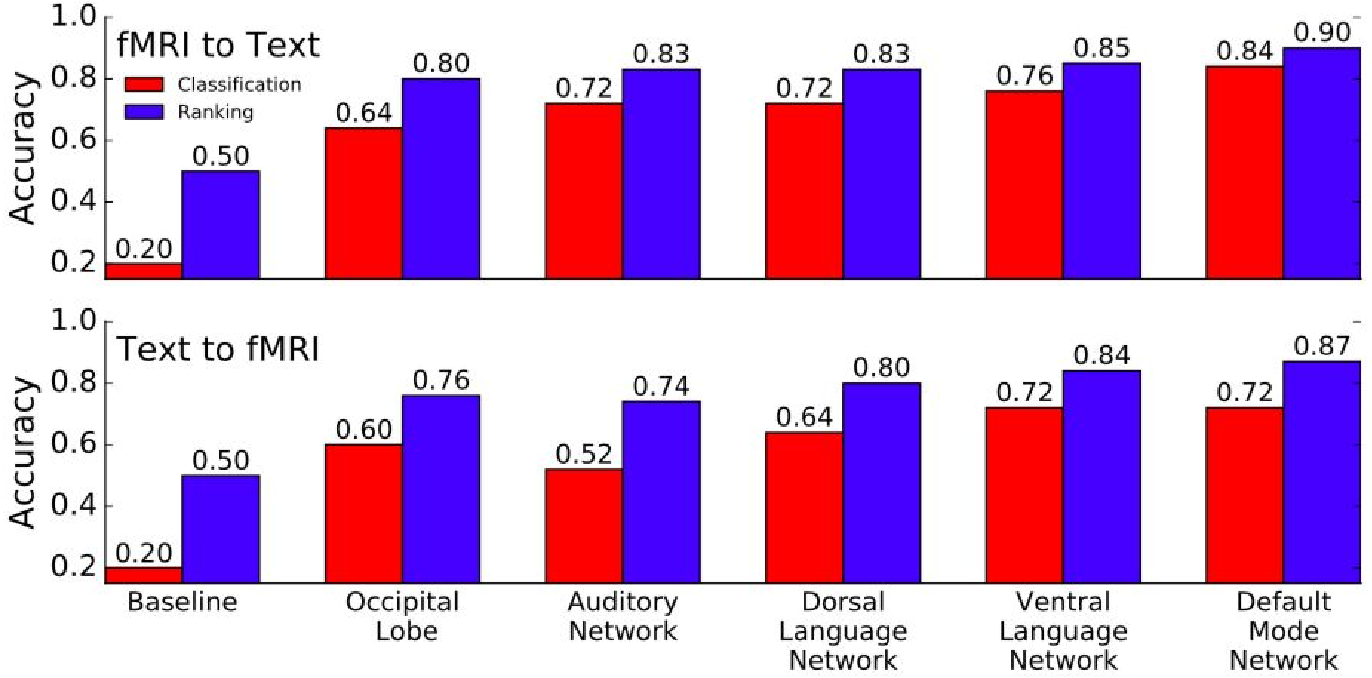


Figure 2: Best Bidirectional Accuracy Scores for Each Brain Region of Interest for both Scene Classification and Ranking (std. err. over different average subsets < 0.01)

Results - Multiplicative Improvement Table

Comparison on the Classification Task	fMRI → Text	Text → fMRI
20-dim SRM / Avg	1.57 ± 0.10	1.00 ± 0.03
Weighted / Unweighted Semantic Vectors	1.17 ± 0.04	1.06 ± 0.03
Temporal Zero Mean / No Zero Mean	1.09 ± 0.04	1.57 ± 0.11
Procrustes / Ridge	1.42 ± 0.09	0.85 ± 0.06

Table 1: Average Improvement Ratio for Various Comparisons

Results - Performance of All Variants on DMN Region

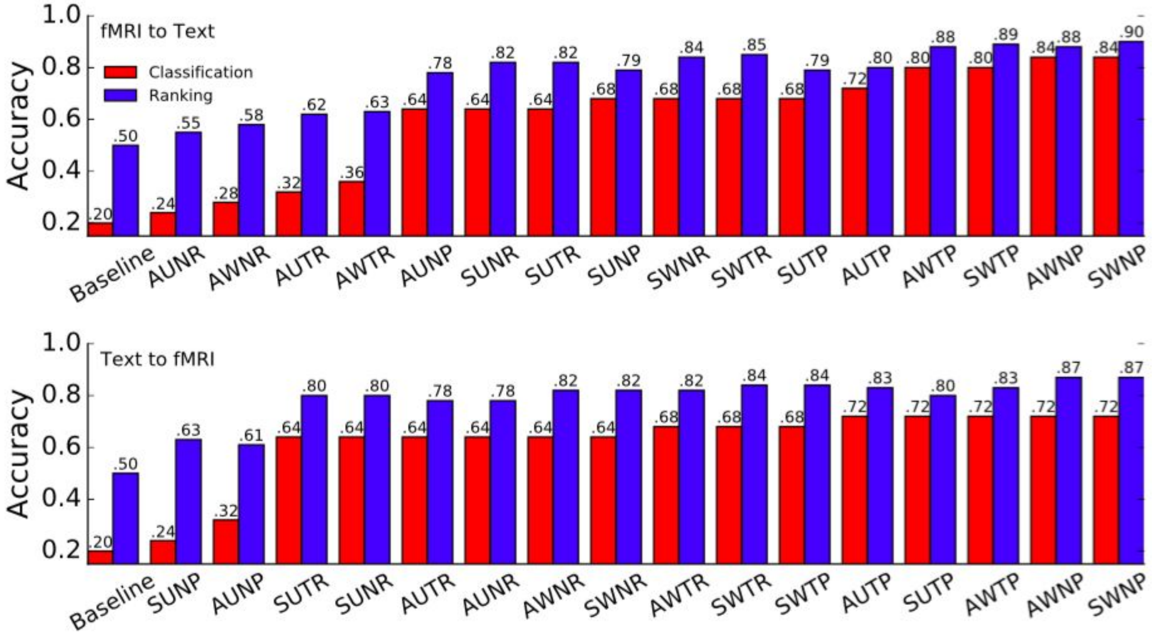


Figure 3: DMN Bidirectional Accuracy Scores for Scene Classification and Ranking. The acronyms stand for combinations of methods, with the following key: S/A = SRM/Average, W/U = Weighting/No Weighted, T/N = Temporal Zero Mean/No Temporal Zero Mean, P/R = Procrustes/Ridge (std. err. over different average subsets < 0.01)