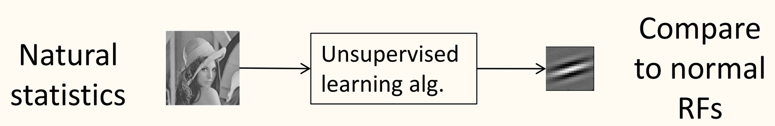
Modeling Cortical Representational Plasticity With Unsupervised Feature Learning

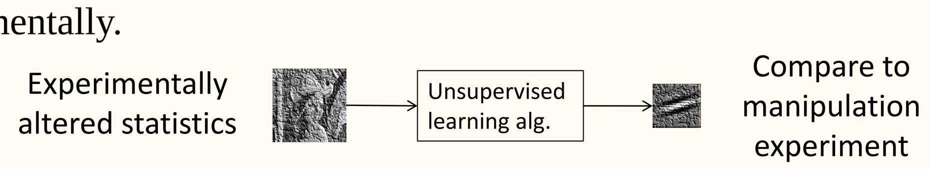
Andrew M. Saxe, Maneesh Bhand, Ritvik Mudur, Bipin Suresh, and Andrew Y. Ng
Stanford University

Overview

The efficient coding hypothesis holds that neural receptive fields are adapted to the statistics of the environment, but is agnostic to the timescale of adaptation. In this work we consider the possibility that neural receptive fields are adapted to the statistics during an organism's lifetime. In particular, we test whether some shared plasticity mechanism can account for normal receptive field properties across multiple primary sensory cortices.



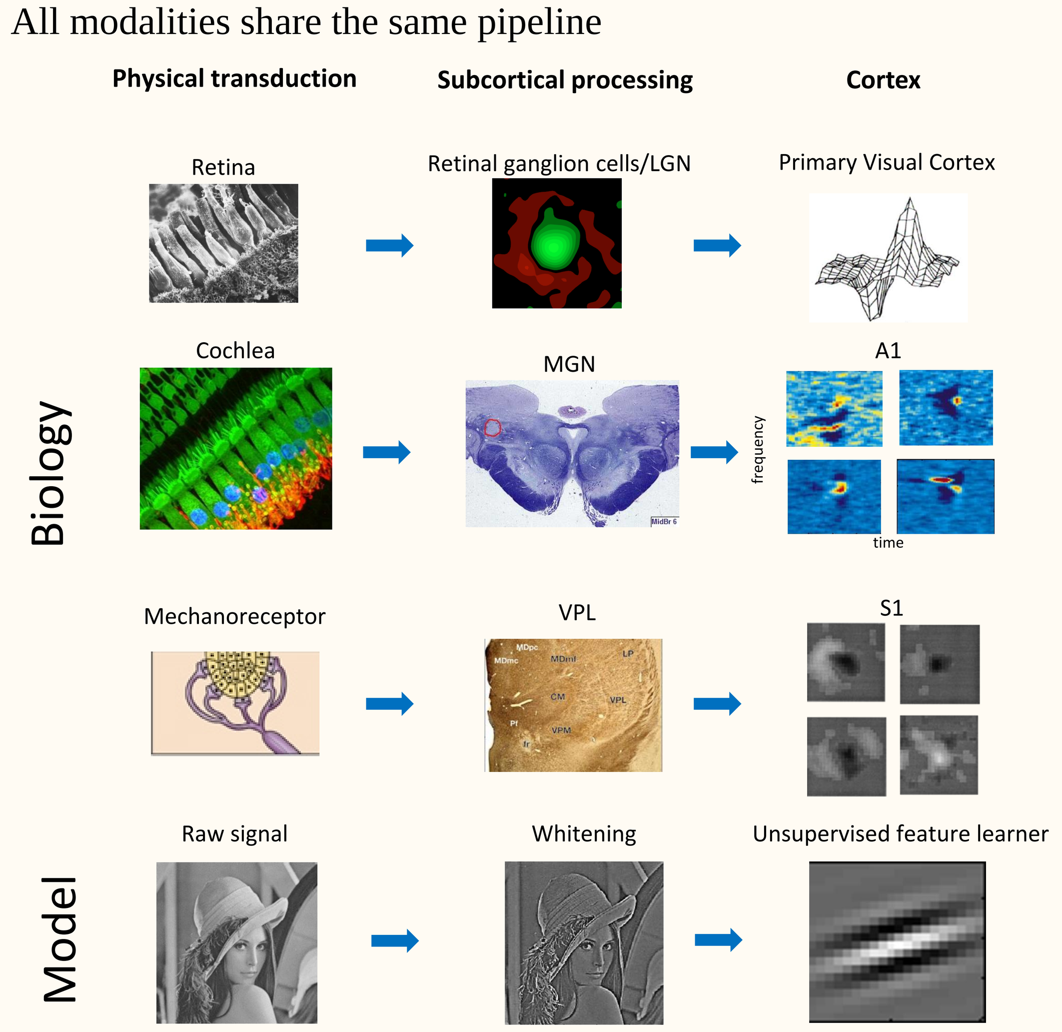
Furthermore, we test whether the same mechanism can account for altered receptive field properties when the statistics of the environment are altered experimentally.



We find that unsupervised feature learning algorithms are able to capture several receptive field properties across sensory modalities, and also allow us to model receptive field plasticity experiments. The consistent correspondences and discrepancies between these algorithms and experimental data may provide insight into plasticity mechanisms and aid theoretical efforts to develop new learning algorithms.

Modeling Approach

- Evaluate learning algorithms on three modalities
- All modalities share the same pipeline



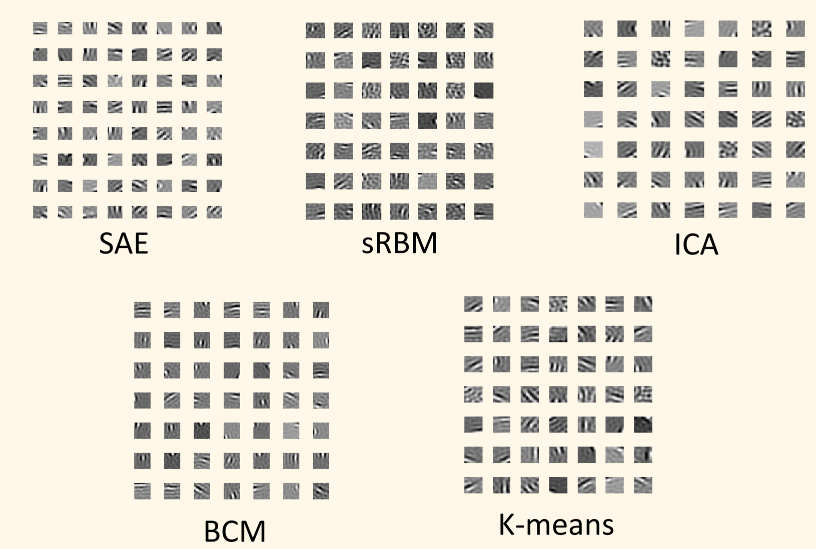
- Compare learned representations to biological data

- We consider five candidate algorithms
  - Independent component analysis
  - Sparse autoencoders
  - Sparse Restricted Boltzmann Machines
  - BCM learning rule
  - K-means

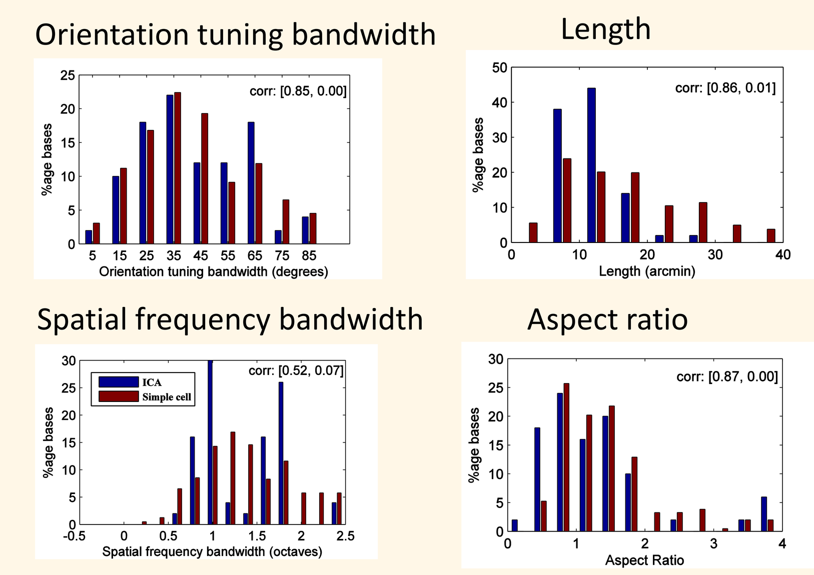
Primary Visual Cortex

- Models trained on natural image data
- Learned representations were localized, band pass filters

Model Receptive Fields

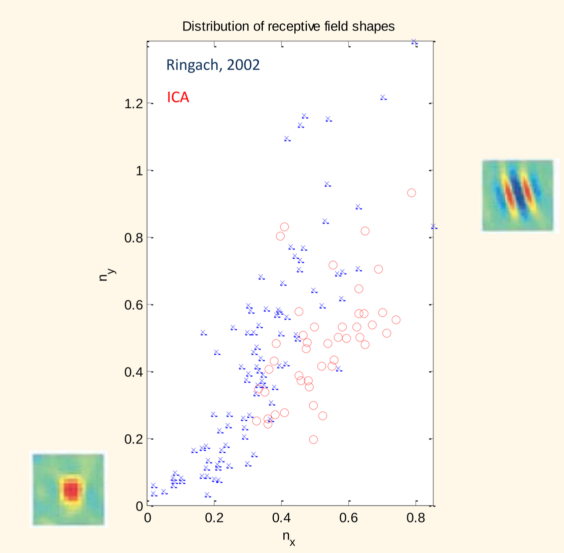


Selectivity Properties



van Hateren and van der Schaaf, 1998

Receptive Field Shape

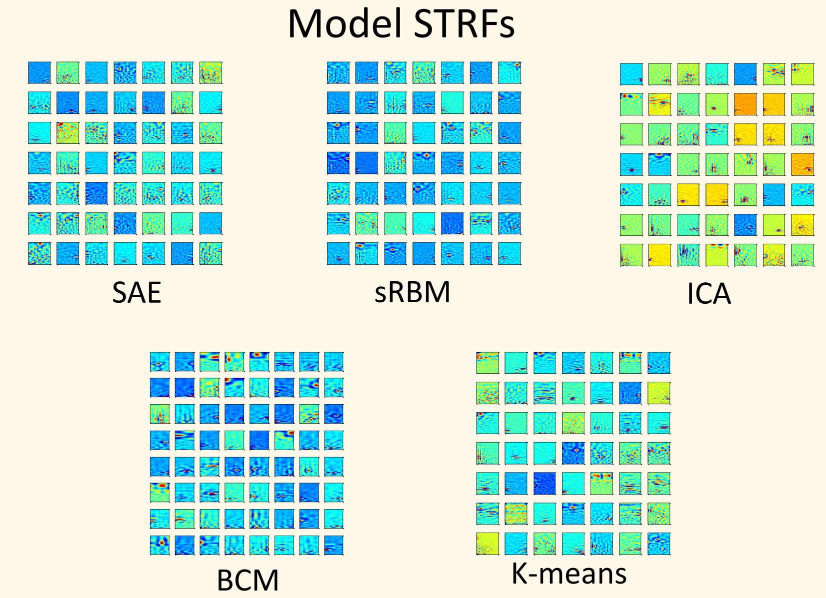


- All models failed to capture low frequency bases

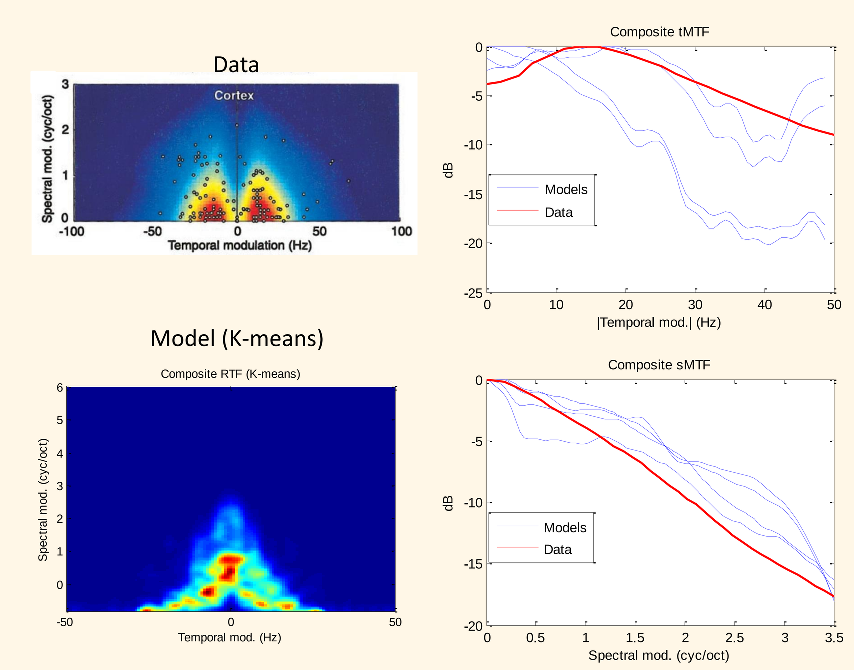
Primary Auditory Cortex

- Models trained on natural sounds and speech data

Model STRFs

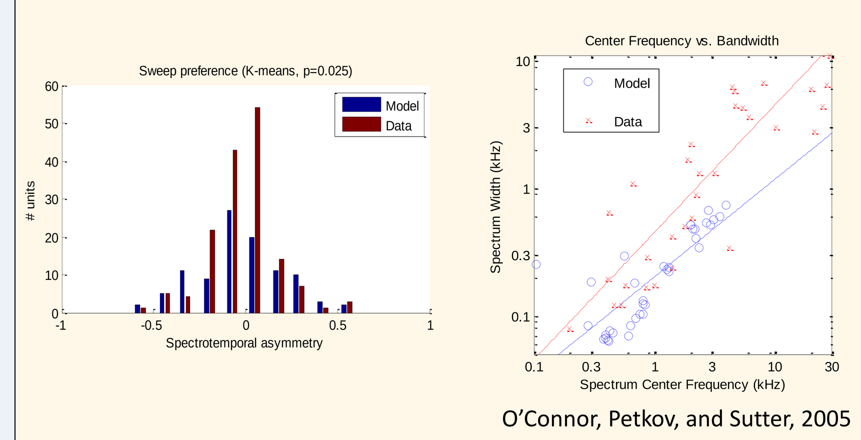


Cortical Ripple Transfer Functions



Miller, Escabi, Read, and Schreiner, 2002

STRF Comparisons



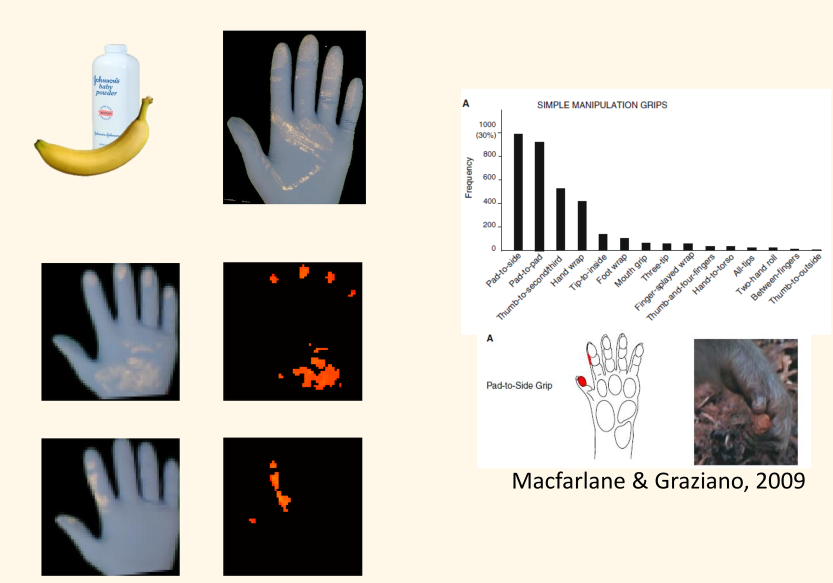
O'Connor, Petkov, and Sutter, 2005

- Model STRFs lack temporal band-pass structure but match spectral low-pass structure
- All models show symmetric sweep preference
- All models show scaling of spectrum width with center frequency

Primary Somatosensory Cortex

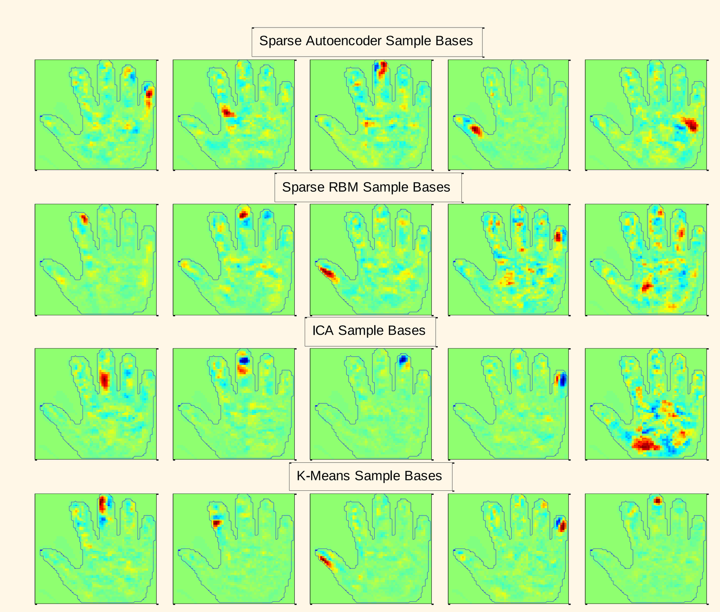
Naturalistic Somatosensory Dataset

- Collected grasp contact patterns with grip types matched to ecological proportions



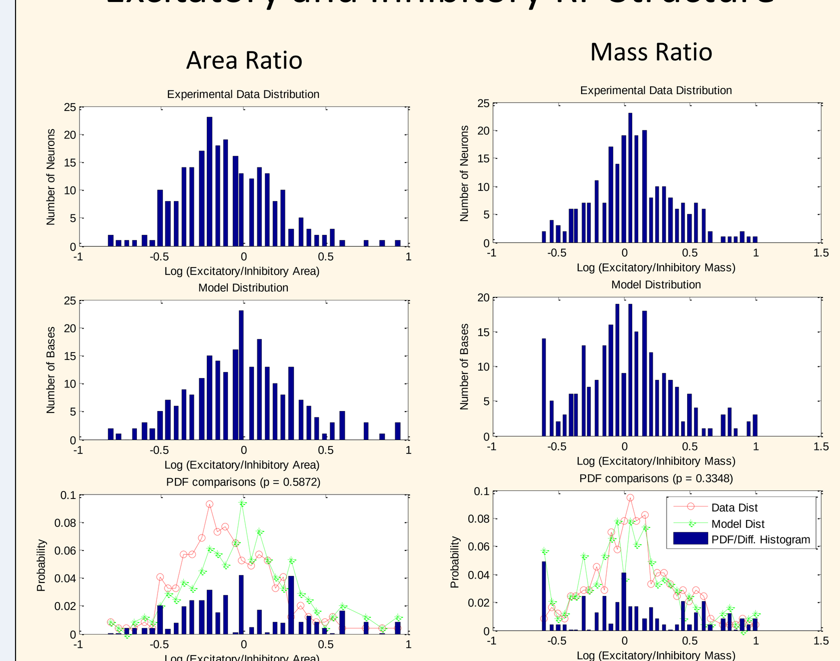
Macfarlane & Graziano, 2009

Model Receptive Fields



- All models develop localized, single digit receptive fields

Excitatory and Inhibitory RF Structure

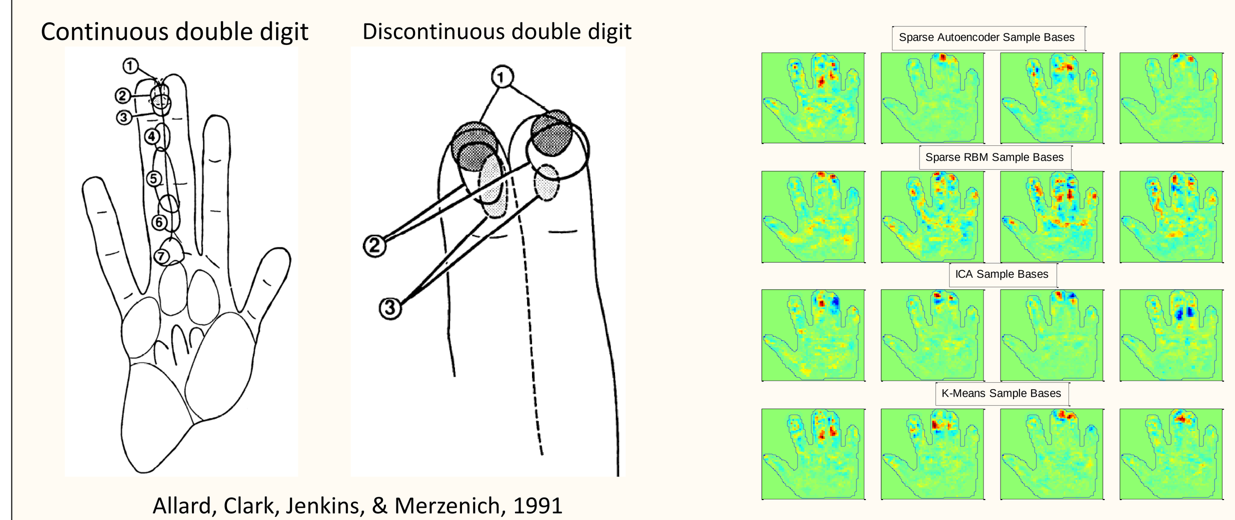


DiCarlo, Johnson, and Hsiao, 1998

- All models match fingertip RF properties in macaque

Plasticity Following Digital Syndactyly

- Surgically fusing digits 3 and 4 forms double digit receptive fields in S1
- We collected naturalistic grip data with fused digits

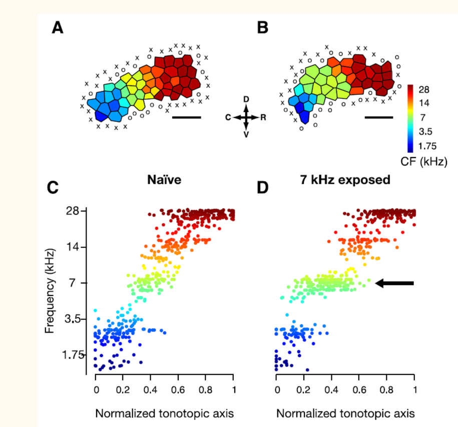


Allard, Clark, Jenkins, & Merzenich, 1991

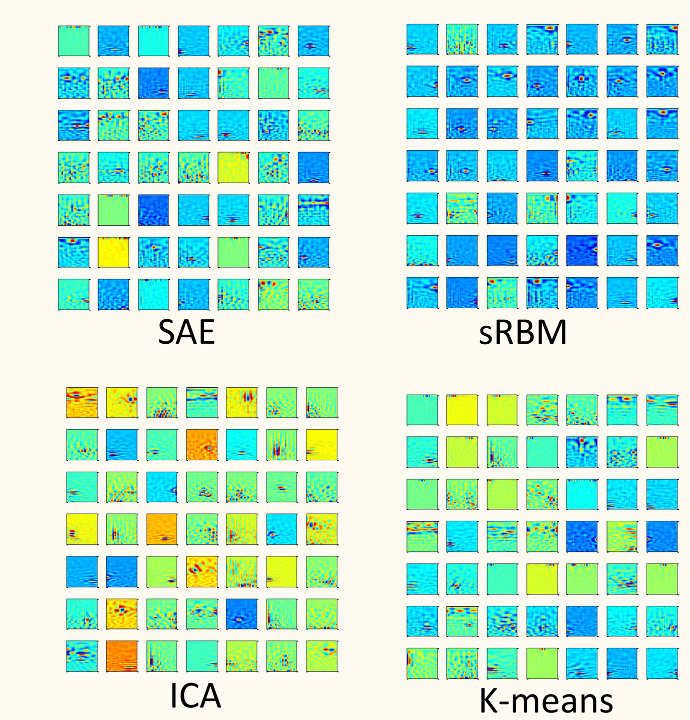
- Model RFs are predominantly two digit receptive fields on digits 3 and 4

Plasticity Following Pulsed-Tone Rearing

- Continuous exposure to a 7kHz pulsed tone train during rearing induces reorganization in rat A1



De Villers-Sidani, Chang, Bao, and Merzenich, 2007.



- K-means and SAE show greater energy in the tone band, whereas ICA and sRBMs show less

Conclusion

- All five of the algorithms yield similar fits to normal receptive field properties in three primary sensory cortices
- Notable discrepancies include a failure to capture low frequency bases in V1, and a marked failure to capture band-pass temporal structure in A1
- Some algorithms accounted for receptive field plasticity following experimental manipulations
- Performance is similar for natural inputs; experiments reveal distinctions between algorithms
- The ability of a simple learning mechanism like K-means to model normal and altered receptive fields suggests that the underlying principles of plasticity may be common across these modalities and phenomena

Acknowledgements: The authors would like to thank Chris Piech, Conal Sathi, and Bill Newsome for code and ideas. This work was supported by an SGF and NDSEG fellowship to A.S.