The Relation Between Activation and Computation in Functional Neuroimaging

Michael C. Mozer
Institute of Cognitive Science and Department of Computer Science
University of Colorado

Martha Farah
Center for Cognitive Neuroscience and Department of Psychology
University of Pennsylvania

Todd Mytkowicz
Department of Computer Science
Colorado State University

Summary

Neuroimaging rests on two assumptions.

- Strength of the image signal (e.g., BOLD) corresponds to aggregate neural activity in a cortical region.
- Aggregate neural activity reflects the effective contribution of that region to cognition.

Although these assumptions are generally supported, exceptions are found.

- A COMT gene polymorphism associated with *lower* prefrontal activation is also associated with *better* executive function.
- *Normal* fusiform face-area activation is found with *impaired* face recognition in congenital prosopagnosia.
- *Decreased* activation throughout cortex with *increased* stimulus-specific experience.

We present a computational framework that provides an interpretation of the exceptions.

Using neural network models, we show how various aspects of the models—such as extraneous noise, fraction of neurons active, amount of training and mere exposure, and the shape of the activation function—can yield dissociations between activity of a brain region and its effective contribution to cognition.

Neuroimaging Assumptions

1. Strength of the image signal (e.g., BOLD) corresponds to aggregate neural activity in a brain region.

Correlations between hemodynamic and neural activity measured in monkeys are generally supportive of this assumption (Disbrow et al., 2000).

2. Aggregate neural activity reflects the contribution of the region to cognitive function.

Exceptions to this assumption can be found, and are puzzling.

Dissociations Between Aggregate Neural Activity and Cognitive Function

The COMT gene polymorphism associated with *lower* prefrontal activation is also associated with *better* executive function performance (Egan et al., 2001)

***Normal* fusiform face-area activation is found with *impaired* face recognition in congenital prosopagnosia (Behrmann et al., 2002)**

***Decreased* activation throughout cortex is observed with *increased* stimulus-specific experience (e.g., Desimone, 1996; Ringo, 1996; Wiggs & Martin, 1998).**

Factors Influencing Aggregate Neural Activity

1. Noise
2. Fraction of neurons active
3. Repeated exposure
4. Domain-based learning

References

Behrmann, M., Marotta, J., Harel, M., & Hasson, U. (2002). Activation in fusiform gyrus is not correlated with face recognition: normal cortical activation with impaired face recognition in congenital prosopagnosia [Abstract]. *Journal of Vision*, *2*, 562a, `http://journalofvision.org/2/7/562/`.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Science*, *93*, 13494–13499.

Disbrow, E. A., Slutsky, D. A., Roberts, T. P. L., Krubitzer, L. A. (2000). Function MRI at 1.5 tesla: A comparison of the blood oxygenation level-dependent signal and electrophysiology. *Proceedings of the National Academy of Sciences*, *97*, 9718–9723.

Egan, M. F., et al. (2001). Effect of COMT Val$^{108/158}$ Met genotype on frontal lobe function and risk for schizophrenia. *Proceedings of the National Academy of Sciences, 98*, 6917–6922.

Jueptner, M., & Weiller, C. (1995). Review: does measurement of regional cerebral blood flow reflect synaptic activity? Implications for PET and fMRI. *Neuroimage*, *2*, 148–156.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*, 150–157.

Nowlan, S. J., & Hinton, G. E. (1993). A soft decision-directed LMS algorithm for blind equalization. *IEEE Transactions on Communications*, *41*, 275–279.

Qureshi, S. U. H. (1985). Adaptive equalization. *Proceedings of the IEEE*, *73*, 1349–1387.

Ringo, J. L. (1996). Stimulus specific adaptation in inferior temporal and medial temporal cortex of the monkey. *Behavioral Brain Research*, *40*, 191–197.

Widrow, B., & Stearns, S. D. (1985). *Adaptive signal processing.* Englewood Cliffs, NJ: Prentice Hall.

Noise

Consider a brain region as transmitting or transforming a signal embedded in noise.

Noise reflects extraneous signals from other brain regions that aren't suppressed.

Aggregate neural activation reflects signal + noise.

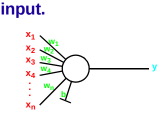
fMRI imaging reflects incoming (and intracortical) neural activity (Jueptner & Weiller, 1995; Logothetis et al., 2001).

Cognitive efficacy depends on signal:noise ratio.

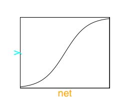
When noise is significant, aggregate neural activation will not correspond to effective contribution of brain region to cognition.

Fraction of Neurons Active

Consider a neuron whose firing rate y is a sigmoidal function of its input.



$$\text{net} = \sum_i w_i x_i + b$$

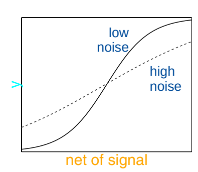


Introducing noise to input reduces the expected response range.

Neurons with strong responses fire less; neurons with weak responses fire more.

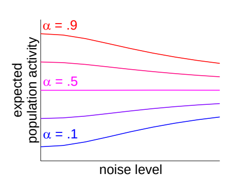
True for any symmetric noise distribution, e.g., Gaussian, uniform

True for other response functions, e.g., piecewise linear with min, max values



Consider population of neurons with some fraction (α) firing high and the remainder firing low.

Interaction between α and noise level



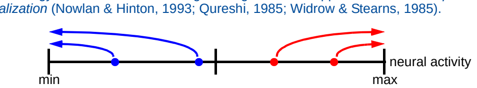
Repeated Exposure

We propose an unsupervised learning rule for adjusting neural net weights as a result of repeated exposure (i.e., experience, practice, repetition priming).

In the absence of specific training signal, a good strategy for neurons is to push their activation to extremes—fully on or fully off.

This strategy makes neural responses more noise robust and is consistent with neurophysiological recording data (Mozer & Mytkowicz, 2004).

This strategy is a standard trick in engineering for noise suppression called *adaptive equalization* (Nowlan & Hinton, 1993; Qureshi, 1985; Widrow & Stearns, 1985).

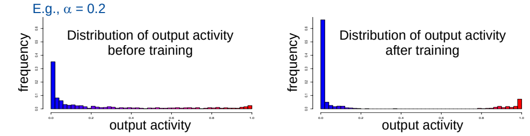


Simulation

A set of input patterns is presented repeatedly to a network, and the adaptive equalization algorithm is applied.

Initial weights are random, and biases are set such that a given fraction (α) of units have high activity.

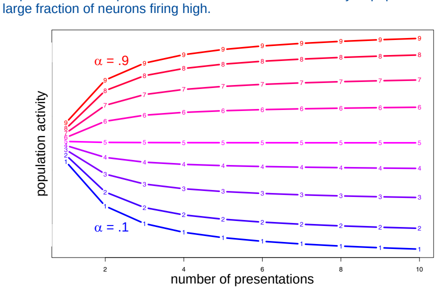
E.g., α = 0.2



Aggregate activity in network depends both on the number of presentations and α.

Repeated stimulus presentation causes increase in activity of populations with a small fraction of neurons firing high.

Repeated stimulus presentation causes decrease in activity of populations with a large fraction of neurons firing high.



Domain-Based Expertise

Certain brain regions are associated with domain-specific activity.

E.g., fusiform face area, parahippocampal place area

Explanation 1

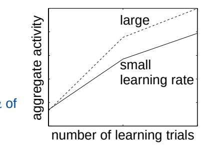
With learning, neurons are recruited to perform domain-specific discriminations.

Consistent with neurophysiological studies in which monkeys receive intense task training: a large fraction of neurons in visual and frontal areas come to discriminate task-relevant features.

Explanation 2

With learning, the amplitude of the output activity grows.

This phenomenon is observed in a neural network whose weights are initially small and whose target output representations have a significant fraction α of active units.



Both explanations suggest increasing aggregate neural activation with learning in domain-specific cortical regions.

Explaining Data

Congenital prosopagnosics show normal FFA response but poor behavioral performance.

noise, in conjunction with a representation characterized by α ≅ 0.5

Repetition suppression effects

repeated exposure, in conjunction with a representation characterized by α < 0.5

Predictions from this framework require assumptions concerning the fraction of active neurons in a brain region.

Implications

In cognitive neuroscience, measures of brain activation are often regarded as primitives that do not require interpretation.

Although activity in a brain region may commonly indicate the effective contribution of that brain region to cognition, we argue on computational grounds that dissociations are possible, and are consistent with experimental data.

Factors that influence aggregate neural activity in computational models include: extraneous noise, fraction of neurons active, amount of training, number of exposures, and the nature of the function relating dendritic input to firing rate.

We welcome pointers to puzzling phenomena that may be explained within this framework, as well as data that could falsify the framework.