Project Title: **Detection and Recognition of Objects in Visual Cortex**
In response to program announcement for Silvio O. Conte Feasibility Centers for Neuroscience Research
(PA number: PAR-02-123)

Universities: Massachusetts Institute of Technology (plus California Institute of Technology and Northwestern University and Georgetown University)

Investigator(s): Tomaso Poggio*, Christof Koch, David Ferster, James Dicarlo, Earl Miller, Maximilian Riesenhuber

Website: http://cbcl.mit.edu/projects/index-projects-NIH-Conte.htm

A succinct description of the Center that captures its current scientific focus, activities and potential for the future.

A recent description of our work (reports and abstracts of the recent meeting) with up-to-date publications is available on the Web, see http://cbcl.mit.edu/projects/index-projects-NIH-Conte.htm

The central focus of our Center is to develop a framework for studying the neural computations underlying object recognition in visual cortex. The Center's framework is based on the collaboration of labs working on monkey physiology, cat physiology and human psychophysics with a quantitative computational model providing the main conduit through which experimental results in one lab affect experiments in another lab. The model represents a novel tool for driving a collaborative enterprise, providing a way to integrate the data, to check their consistency, to suggest new experiments and to interpret the results. The theory itself, based on two existing models for recognition and attentional saliency, not only guides the experiments and is a conduit for synergies between different labs but is also evolving and improving as an effect of the experimental results.

The research is organized into three main projects, defined by geographical location and scientific questions, rather than discipline. In the MIT project, the labs of Tomaso Poggio, Earl Miller and James DiCarlo are guided by a quantitative hierarchical model of recognition, probing the relations between identification and categorization and the properties of selectivity and invariance of the neural mechanisms in IT and PFC cortex. In the Northwestern project, the lab of David Ferster is testing some key aspects of the model about the nature of the pooling operation (a *max* operation vs. a linear sum) performed by complex cells in V1 (and in other cortical areas such as V4) using very similar stimuli affected by clutter; they will also test the mechanisms underlying tuning of visual neurons. In the Caltech project, the lab of Christof Koch will collaborate with Ferster lab on biophysical simulations of V1 circuits. It is also testing -- using human psychophysics with some of the same stimuli used by Jim DiCarlo -- the conditions under which *attention* is needed in recognition; from the data, in collaboration with Tomaso Poggio, his lab is working on extending the basic model of recognition by integrating it with a *saliency* model.

A unique aspect of our Center is that, the computational component, centered around a quantitative model of recognition, is the generic tool to drive and manage interactions between the investigators, in addition to the standard pair-wise interactions: the model suggests an experiment and guides its planning and interpretation; the experimental results from one lab impact, through the model, work done in another lab, including model development, as well as their interpretation and what to do next. Ultimately, the whole process should lead to a better and more coherent understanding of the neural mechanisms of visual recognition. Thus the Center's key motivations is the belief that quantitative models of complex neural system, when developed in close cooperation with experimental labs, can be tools to a) think about the problems (some cognitive problems are too complex for the qualitative, simple models used so far); b) make predictions, suggest and plan new experiments; c) analyze and interpret data; d) integrate experimental findings of different types and from different labs, drawing implications for future work from multiple sources of evidence.

1-3 recent findings (with citations) that you regard as your Center's most salient.

We would like to mention several recent findings:

- The first one regards plausible neural circuits underlying tuning in neurons in higher visual cortical areas. An interesting answer from ongoing modeling work is that Gaussian-like, multidimensional tuning can be generated by *normalization* of the input vector, followed by a simple threshold-like sigmoidal nonlinearity. A specific circuit for normalization, based on lateral shunting inhibition, is described in a very recent paper (together with general arguments for Gaussian-like tuning in cortex), see Poggio, T. and E. Bizzi. Generalization in Vision and Motor Control, *Nature*, Vol. 431, 768-774, 2004. In the past, various neurally local circuits have been proposed to implement the key normalization stage, though the motivation was to account for gain control and not tuning properties. We make here the new proposal that another *role for normalizing local circuits in the brain is to provide the key step for neural tuning in cortex and therefore for the key ability to generalize.* In fact this may be the main reason for the widespread presence of gain control circuits in cortex where tuning to optimal stimuli is a common property. Thus, our new hypothesis is that *gain control microcircuits underlie tuning of cells* to optimal stimuli in both the visual and motor system.

- A second interesting finding is that a the standard model of visual cortex that we have been developing provides the basis for a computer vision system that outperforms the best existing object recognition systems on a variety of databases for many different object classes. This is interesting for computer vision and suggests a natural extension of the standard model on which we are working (see Fig. 1).

- Thirdly, Miller's lab found strong differences in PFC neurons (but not in ITC neurons) between the category task vs passive viewing. These differences were not found in a categorization vs an identification task (see Figs 2,3,4).

- Fourthly,on the experimental side, a collaborator of Ferster's lab, Ilan Lampl at the Weizmann Institute in Rehovot/Israel, has carried out intracellular recordings in the rodent barrel cortex while stimulating one or two whiskers. Remarkably, his data indicates the existence of a MAX like computation, similar to what he had found earlier on as a post-doctoral scholar in Ferster's lab. This suggests that a MAX like computation may be a canonical operation thoughout cortical regions and not only restricted to vision.

- Finally, the Caltech team demonstrated that a combination of saliency-based visual attention and object recognition leads to superior recognition performance in cluttered environment. The poster presenting this research was given the "best poster award" at the Computer Vision and Pattern Recognition 2004 conference (Rutishauser U, Walther, D, Koch C and Perona P "Is attention useful for object recognition" IEEE Intl. Conf. on Computer Vision and Pattern Recognition, II, 37-44, 2004 and Dirk Walther, Ueli Rutishauser, Christof Koch, and Pietro Perona, Selective visual attention enables learning and recognition of multiple objects in cluttered scenes, Computer Vision and Image Understanding, in press).

Fig.1: *Extensions of the model of visual recognition (Serre and Poggio, ongoing work; see Riesenhuber and Poggio, Nature Neuroscience 2000)*

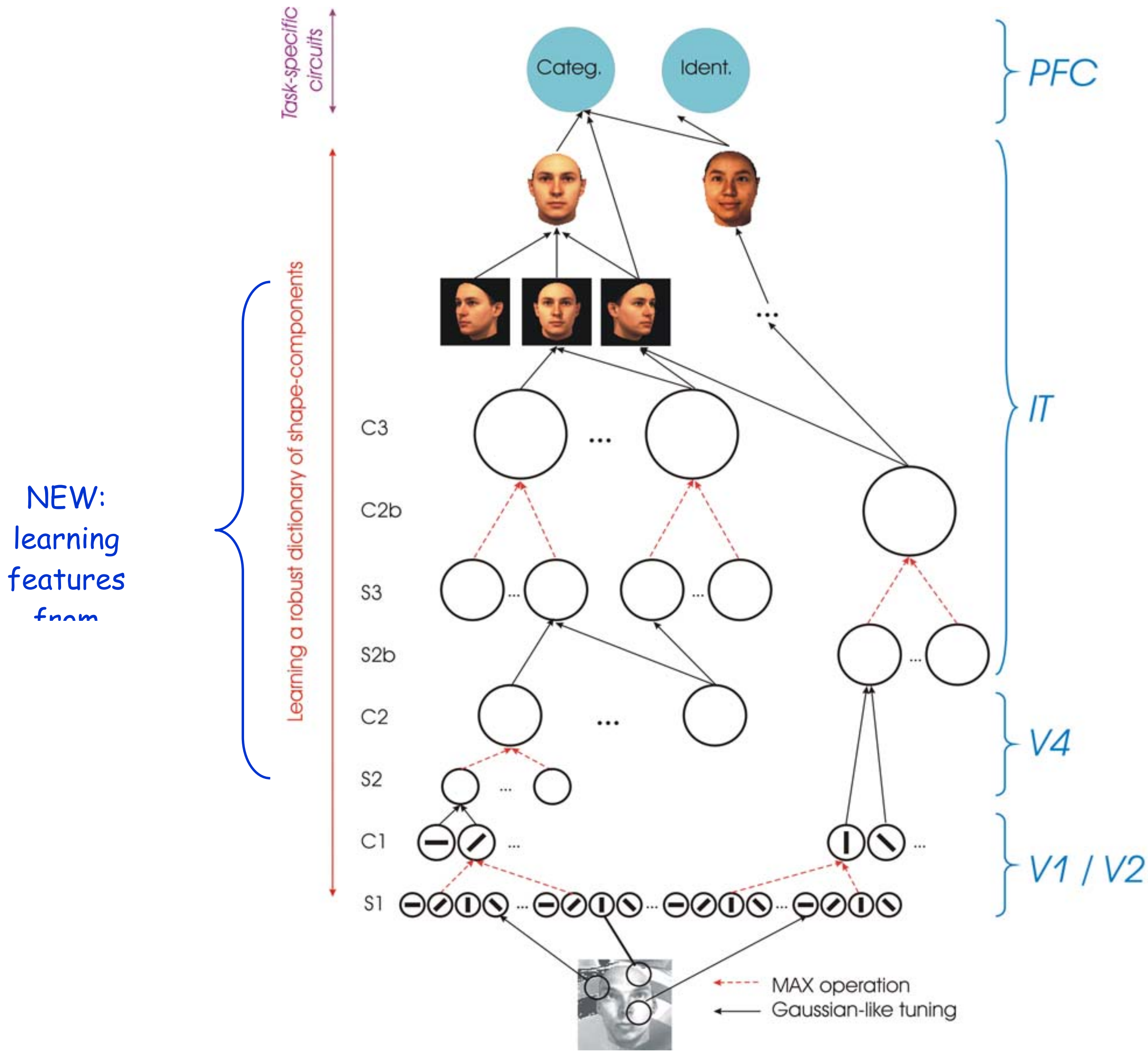
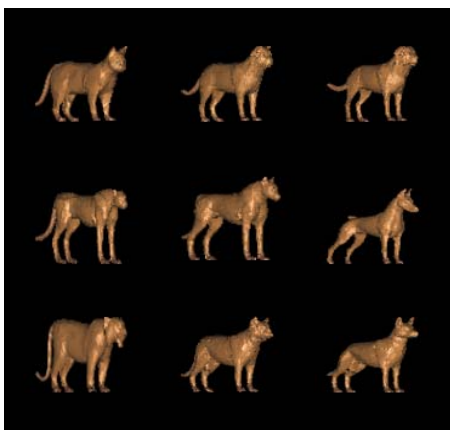


Fig.2,3,4: *There are strong differences in PFC neurons (but not in ITC neurons) between the category task vs. passive viewing. These differences were not found in a categorization vs an identification task (see Figs 2,3,4).*

Are stimulus representations in PFC and ITC modulated by changes in task demands?

Recorded from neurons in PFC and ITC while monkeys alternated between categorization task and passive viewing.



DMC task → passive viewing …

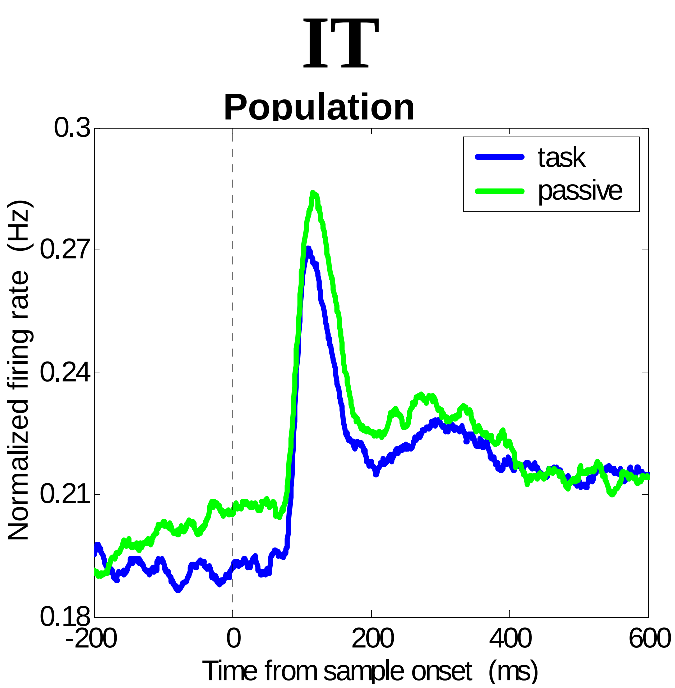
100 trials 200 trials

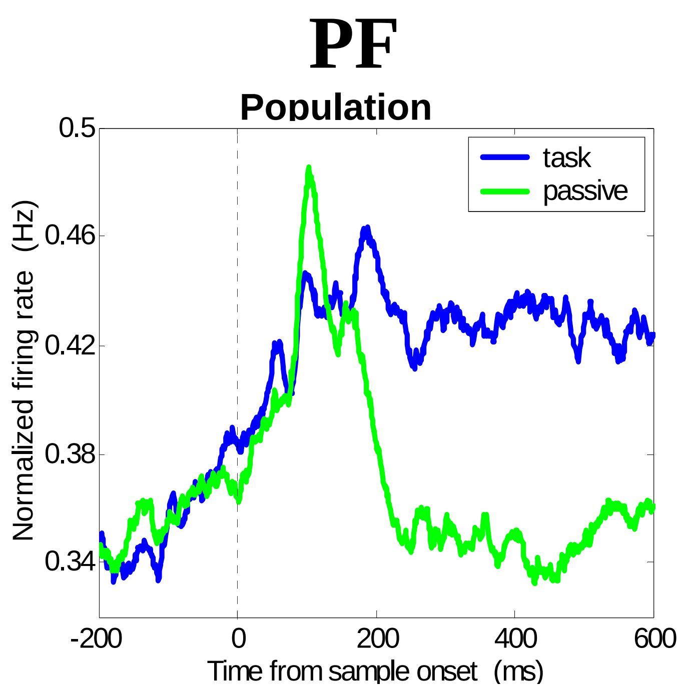
cats and dogs (n=42)

TASK + PASSIVE

Simultaneous recordings from 298 ITC, 212 PFC neurons from two monkeys

Comparison of firing rates to cats/dogs during task and passive viewing





ITC activity similar between task and passive viewing.
PFC responses were more task-dependent.

How was category selectivity modulated by task demands?