

## Abstracts of NIH-Conte Meeting – August 29, 2005

### The Relationship of Spatial Frequency Tuning to the Substructure of Receptive Fields in Complex Cells of Cat Primary Visual Cortex

Ian Finn and David Ferster

*Northwestern University*

The receptive fields (RF) of complex cells are uniform in their responses to single flashed bars, a property that predicts broad tuning for stimulus spatial frequency (SF). In many cells, however, interactions between paired stimuli reveal a substructure to the receptive field, with facilitation at small interstimulus distances and suppression at larger distances. For example, the response to a single fixed bar is greater when a bar of equal polarity is presented close by, and less when the second bar is placed further away. This substructure resembles the receptive fields of simple cells, and, compared to the overall RF structure, predicts much sharper spatial frequency tuning that more closely matches the tuning measured with drifting gratings (Movshon *et al.* 1978, *J. Physiol* 283:79). In intracellular recordings, we have found a number of cells that behave as these authors described. Other cells, however, were more MAX-like in that their responses to pairs of bar stimuli were comparable to the largest of the two responses evoked when the bars were presented individually (Lampel *et al.* 2004). This MAX-like behavior extended across the entire receptive field, and was observed for all relative separation distances between the bars. As predicted by the absence of substructure, SF tuning in these cells was much broader than in those cells with simple-like stimulus interactions. To understand how complex cell substructure might generate spatial frequency selectivity, we have applied principal component analysis (Touryan *et al.* 2005, *Neuron* 45:781; Rust *et al.* 2004, *Neurocomputing* 58-60:793) to the intracellularly-recorded membrane potential responses of cells stimulated with one dimensional noise. In those cells possessing clear receptive field substructure, the underlying spatio-temporal principal components recovered from the analysis accounted well for the spatial frequency tuning and direction selectivity of the cells when measured with drifting gratings.

### Multiple Object Response Normalization in Monkey Inferotemporal Cortex

Davide Zoccolan

*Massachusetts Institute of Technology*

The highest stages of the visual ventral pathway are commonly assumed to provide robust representation of object identity by disregarding confounding factors such as object position, size, illumination and the presence of other objects (clutter). However, while neuronal responses in monkey inferotemporal cortex (IT) can show robust tolerance to position and size changes, previous work shows that responses to preferred objects are usually reduced by the presence of non-preferred objects. More broadly, we do not yet understand multiple object representation in IT. In this study, we systematically examined IT responses to pairs and triplets of objects in three passively viewing monkeys across a broad range of object effectiveness. We found that, at least under these limited clutter conditions, a large fraction of each IT neuron's response to multiple objects is reliably predicted as the average of its responses to the constituent objects in isolation. That is, multiple object responses depend largely on the relative effectiveness of the constituent objects, regardless of object identity. This average effect becomes virtually perfect when populations of IT neurons are pooled. Furthermore, the average effect cannot simply be explained by attentional shifts, but behaves as a largely feed-forward response property. Taken together, our observations are most consistent with mechanistic models in which IT neuronal outputs are normalized by summed synaptic drive into IT or spiking activity within IT, and suggest that normalization mechanisms previously revealed at earlier visual areas are operating throughout the ventral visual stream.

## **A Theory of the Feedforward Path of the Ventral Stream in Visual Cortex: Human Level Recognition Performance?**

**Thomas Serre**

*Massachusetts Institute of Technology*

During this talk I shall describe a new biologically plausible trace-learning rule to generate a dictionary of shape-tuned units from V4 to IT from the model passive exposure to many natural images.

Extensive tests on large-scale real-world object recognition tasks show that learning improves recognition performance drastically. Indeed, the model competes and even outperforms state-of-the-art computer vision systems. Experimental results suggest that the tuning properties of the model C2 units generated after learning is congruent with data collected in V4 (Pasupathy and Connors, 2001). The most significant result is that the model performs at human level on an ultra-rapid (animal/non-animal) categorization task of the type used by Simon Thorpe, for which one expects that back projections do not play any significant dynamic role.

Because of its ability 1) to generate shape-tuned units compatible with neural data and 2) to perform recognition at the level of humans, the model is a unique tool to make robust non-trivial predictions.

## **A Theory of the Feedforward Path of the Ventral Stream in Visual Cortex: IT**

**Minjoon Kouh**

*Massachusetts Institute of Technology*

Based on many physiological findings that IT neurons are selective to images of objects (faces, hands, trained objects, etc.) with some degree of invariance, the analogous model units are constructed to be selective to particular activation patterns of their afferent neurons with translation and scale invariance properties. In this short talk, we show that a gain-controlled weighted sum can give rise to a tuning behavior, just like the Gaussian function in the original version of the model. We compare with several experimental results and verify that, using such tuning operations and a feedforward hierarchy of increasing invariance and selectivity, the model IT neurons yield similar range of selectivity/invariance against distractors on paperclip stimuli, similar average-like behavior under clutter conditions, and similar information capacity to allow read-out of the current stimuli (preliminary results).

## **A Theory of the Feedforward Path of the Ventral Stream in Visual Cortex: V4**

**Charles Cadieu**

*Massachusetts Institute of Technology*

We show that a quantitative model of visual processing – originally developed to explain tuning properties of some IT neurons – is capable of reproducing, predicting, and explaining the responses of neurons in area V4. Using optimization techniques we have developed, model neurons accurately predict the responses of 8 V4 neurons to within-class stimuli, such as closed contours (recordings by Pasupathy and Connor) and gratings (recordings by Freiwald, Tsao and Livingstone), and achieve an average correlation coefficient of 0.77 between predicted responses and measured V4 responses. In addition, by fitting the model neurons to a V4 neuron's grating response, it is possible to qualitatively predict the V4 neuron's 2-spot reverse correlation map. Upon further analysis, we propose that V4 2-spot reverse correlation maps are explained by relative position coding in V4 and not by orientation selectivity of V1-like subunit afferents. By reproducing, predicting, and explaining V4 representation with a nonlinear combination of V1 neural responses, our results bridge V1 and V4 experimental data.

We thank Anitha Pasupathy, and Charles Connor, and Winrich Freiwald, Doris Tsao, and Margaret Livingstone for providing V4 recording data and insightful comments.

## A Theory of the Feedforward Path of the Ventral Stream in Visual Cortex: Biophysics

**Ulf Knoblich and Tomaso Poggio**

*Massachusetts Institute of Technology*

Several groups have reported that some neurons in visual cortex respond rapidly and sublinearly to the combined presentation of two simple stimuli in their receptive field [Gawne *et al.* 2002, Lampl *et al.* 2004]. It has been proposed that, instead of the sum, these neurons compute either the maximum [Riesenhuber & Poggio 1999] or the average [Chelazzi *et al.* 1998] of its inputs.

Based on the general architecture of the circuits presented in [Yu *et al.* 2002], we present biophysically plausible models of microcircuits for computing the maximum and average operations. We assume that each input contributes an excitatory as well as an inhibitory component to the output of the circuit. The synapses in our circuits are modeled according to the kinetics described previously [Destexhe *et al.* 2001]. The nonlinear transfer function of the synapses together with their temporal characteristics gives rise to a sublinear response of the circuit, closely matching several aspects of experimental results from [Finn *et al.* as well as Lampl *et al.*]. New experiments derived from these simulations will further help to determine the true nature of the cortical mechanisms involved in the observed sublinear computations.

## Activity of Monkey Prefrontal Cortex Neurons during Shifts in Category Membership

**Jefferson Roy, Maximilian Riesenhuber, Tomaso Poggio and Earl Miller**

*Massachusetts Institute of Technology*

We often categorize items differently depending on their behavioral context. So we recorded neural activity in the lateral prefrontal cortex (PFC) during a task requiring dynamic reallocation of category membership of a set of visual stimuli. Previously, we had reported that PFC neurons reflected the category membership of a continuous set of morph images divided into “cats” versus “dogs.” Here, four prototypes were used to generate a stimulus set that was divided along two different category schemes with orthogonal boundaries. A cue signaled which category scheme should be followed on a given trial (i.e., which boundary was relevant) and the monkey indicated whether two successively presented stimuli were from the same category based on the currently relevant boundary.

We are finding that two different factors affect the activity of PFC neurons: the currently relevant category boundary (the “rule/boundary”) and the category membership it dictated. Preliminary data indicated that 63 PFC neurons showed a rule/boundary and/or category membership effect during the sample and/or the delay epochs (37% of all recorded neurons;  $p < 0.05$ ). These neurons showed activity that varied with the relevant rule/boundary, and often superimposed on this was activity that signaled the category membership of the stimuli, but only for the currently relevant categories. In fact, many individual neurons dynamically switched their selectivity with the relevant category scheme. These results indicate that different category schemes can be “multiplexed” on single neurons and that PFC neurons can dynamically change their category selectivity with behavioral demands.

## When Do You Need to Pay Attention When Recognizing Objects?

**Rufin Van Rullen**

*California Institute of Technology*

A 25-year old tradition in experimental psychology holds that object recognition can only be achieved following the recruitment of attentional processes. Yet feed-forward vision models such as HMAX clearly demonstrate that high-level representations supporting some form of recognition can easily be constructed hierarchically without involving attention. In line with these theoretical results, our psychophysical investigations have established that, at least for well-known familiar and/or natural object categories, focal attention is not a prerequisite for recognition or categorization. So when do you need to pay focal attention? First, our results show that attention is required to recognize unfamiliar and/or synthetic objects, for which the visual system had no time (or no reason) to develop hierarchical feed-forward strategies. Second, attention is required when spatial competition between objects in the visual

field (or other sources of noise) precludes normal functioning of the system: for example in visual search experiments with small inter-stimulus distances, where search performance is not parallel, even for familiar and natural object categories. In sum, the “pre-attentive” human visual system fails precisely in those situations where feed-forward hierarchical neural network models such as HMAX meet their current limits, suggesting that these recognition models are not very far from accurate, and that biologically inspired attentional “feed-back” processes may be the logical next step to improve their performance.

## Modeling Feature-sharing between Object Detection and Top-down Attention

**Dirk Walther**

*California Institute of Technology*

Visual search and other attentionally demanding processes are often guided from the top down when a specific task is given (e.g. Wolfe *et al.* Vision Research 44, 2004). In the simplified stimuli commonly used in visual search experiments, e.g. red and horizontal bars, the selection of potential features that might be biased for is obvious (by design). In a natural setting with real-world objects, the selection of these features is not obvious, and there is some debate which features can be used for top-down guidance, and how a specific task maps to them (Wolfe and Horowitz, Nat. Rev. Neurosci. 2004). Learning to detect objects provides the visual system with an effective set of features suitable for the detection task, and with a mapping from these features to an abstract representation of the object. We suggest a model, in which V4-type features are shared between object detection and top-down attention. As the model familiarizes itself with objects, i.e. it learns to detect them, it acquires a representation for features to solve the detection task. We propose that by cortical feedback connections, top-down processes can re-use these same features to bias attention to locations with higher probability of containing the target object. We propose a model architecture that allows for such processing, and we present a computational implementation of the model that performs visual search in natural scenes for a given object category, e.g. for faces. We compare the performance of our model to pure bottom-up selection as well as to top-down attention using simple features such as hue.

## Shape and Category Tuning in Human Cortex Induced by Categorization Training: A fMRI-RA Study

**Xiong Jiang and Maximilian Riesenhuber**

*Georgetown University*

The knowledge of categories is important for humans to process visual objects efficiently and appropriately based on their category membership. However, the neural mechanisms underlying categorization and category learning, in particular in humans, are still poorly understood. In our previously presented computational model of object recognition in cortex (Riesenhuber & Poggio, 2000), object categorization is mediated through a shape-specific (but category-agnostic) representation of relevant stimuli in inferotemporal cortex (IT) that provides input to task-specific (in this case, category-specific) circuits further downstream, e.g., in prefrontal cortex (PFC). This scheme has subsequently been supported by electrophysiological studies (Freedman *et al.*, 2001, 2002, 2003) in which monkeys were trained on cat/dog categorization tasks, followed by recordings in IT and PFC. We now investigated whether human category learning is based on similar principles. For this purpose, we trained human subjects on a delayed-match-to-category task using morphed images (cars), and investigated their brain activation before and after training using an fMRI rapid adaptation paradigm that allowed us to probe neuronal tuning more directly than previous techniques based on average BOLD contrast response to object of interest. While inside the scanner, two car images were displayed to subjects in each trial, and subjects were asked to judge either motion direction (apparent motion task, pre- and post-training) or category membership (categorization task, post-training only) of these two cars. With the comparison of pre- and post-training for the apparent motion task, we observed a shape-specific change in the ventral stream of visual cortex -- more focused activation pattern and increased selectivity to changes in car shape, but no evidence of category information. By contrast, a strong category-related effect was observed in the inferior frontal cortex while subjects were doing the categorization task (post-training only). However, such a categorization-related activity was much diminished when subjects were doing

the apparent motion task (post-training). These results strongly support the model and suggest that human and monkey category learning follow similar principles.

### **Causal Evidence Linking IT Neural Activity to Object Recognition**

**Chou Hung**

*Massachusetts Institute of Technology*

Our understanding of the role of the ventral stream in object recognition has stemmed from two primary avenues of experiments – disruption via lesions and cooling probes; and correlation between neural activity and object recognition behavior. Causal evidence linking IT neural activity to object recognition behavior has been inconclusive, and it is thought that electrical stimulation of temporal cortex (on the order of several mA of current) leads to conscious recognition only through indirect activation of more medial (e.g. hippocampal) structures.

Here, we hope to demonstrate that activation via electrical *microstimulation* of *small* clusters of neurons in macaque anterior inferotemporal cortex can lead to specific, predictable biases in object recognition behavior. For each electrode site, a pair of objects was chosen based on the selectivity of multiunit responses at that site. We tested the behavioral performance of the monkey during a discrimination/working memory task in which a blend of the two objects was (on random trials) paired with electrical microstimulation. Importantly, response targets were randomized such that the effect of microstimulation on object perception could be dissociated from influences on eye movement planning. Results show a significant positive influence of microstimulation on object recognition behavior in one monkey and are currently being replicated.

### **Fast Decoding of Object Identity**

**Gabriel Kreiman**

*Massachusetts Institute of Technology*

Understanding the brain computations leading to object recognition requires quantitative characterization of the information represented in inferior temporal cortex (IT). We used a biologically plausible, classifier-based read-out technique to investigate the neural coding of selectivity and invariance at the IT population level. We found that the activity of small neuronal populations (~100 randomly selected cells) over very short time intervals (as small as 12.5 ms) contains surprisingly accurate, robust and invariant information about objects. Coarse information about position, scale and stimulus onset can also be read-out from the same population. The approach shows that invariant object information may be rapidly, accurately and robustly decoded by downstream neurons. We will also show data comparing the results from IT neurons against the performance of different stages of the hierarchical model of ventral visual cortex. (Joint work with Chou Hung, Jim DiCarlo, Tommy Poggio and Minjoon Kouh.)

## Abstracts of NIH-Conte Meeting – August 30, 2005

### Amplification Mechanism of Cortical inhibition and its Role in Spatial Integration

**Ilan Lampl**

*Weizmann Institute of Science*

Spatial integration, a process by which the basic elements of a natural stimulus are combined into a single percept, is fundamental to sensory processing. Here, we study the early stages of this process, in the neuronal circuitry that shapes the response of somatosensory neurons to multiple stimuli. In most cortical areas, the neuronal response to multiple stimuli is smaller than the sum of the responses to each of these stimuli presented alone. Using patch recordings, we found that this sublinear behavior is produced by supralinear activation of the inhibitory inputs. A simple feed-forward model, in which a threshold for the response of inhibitory cells plays a critical role, explains this amplification of inhibition and reveals fundamental differences in the response properties of excitatory and inhibitory inputs.

### Categorical Representation of Visual Motion Direction in the Lateral Intraparietal Area

**David Freedman**

*Harvard Medical School*

Categorization is the process by which behavioral significance is assigned to sensory stimuli. While much known about the neural encoding of simple visual features, less is known about how the brain determines the behavioral meaning of stimuli. This study's goal is to understand the progression from rigid sensory encoding to more flexible cognitive representations during visual categorization.

We trained a monkey to group random-dot motion stimuli into two categories. 360° of motion directions were divided by a learned "category boundary". Directions between -45° and 135° were in one category while the remaining directions were in another. The monkey performed a delayed match-to-category (DMC) task and indicated, by releasing a lever, whether a test stimulus was the same category as a previously presented sample. A hallmark of categorization is that similar stimuli of different categories are treated differently while dissimilar items of the same category are treated similarly. This was evident in the monkey's behavior: it performed with >80% accuracy, even for stimuli near the category boundary.

We recorded 92 LIP and 40 MT neurons during DMC task performance. The pattern of direction selectivity differed strikingly between LIP and MT. In LIP, neurons reflected stimulus category: activity was similar for directions in the same category and differed for directions of different categories. A category-tuning index, which measured activity differences across pairs of directions, revealed larger differences for directions in different categories and similar activity for those in the same category (T-Test,  $P < 0.01$ ). In MT, neurons did not group directions by their category. The category-tuning index revealed that MT neurons showed no influence of the boundary ( $P > 0.5$ ).

This suggests that LIP encodes the behavioral relevance of visual motion, while MT provides a more faithful representation of motion direction.

### Single-Cell Mechanisms for Representing Faces in an fMRI-identified Macaque Face Patch

**Doris Tsao**

*Harvard Medical School*

Functional magnetic resonance imaging (fMRI) in humans and more recently in macaques has revealed specialized regions of cortex which show increased activation to faces compared to other visual objects, but the precise function of these regions is strongly debated. A fundamental question is whether these fMRI-identified face-selective regions encode only faces or other objects as well. This issue is especially puzzling in the macaque monkey: the fMRI results indicate a 7-fold higher response to faces compared to

objects within the macaque face patches, but single-unit physiologists have reported a maximum of only 20% of cells in any given region of the temporal lobe as face-selective. One obvious possibility is that previous single-unit recordings did not target the appropriate regions. We have performed single-unit recordings in a region identified in the same animal by fMRI as being face-selective. This region is in the lower bank of the STS in caudal TE, and likely constitutes the macaque homologue of the human fusiform face area (FFA). We find that in this region the firing patterns of all (100%) of the single units recorded so far (N=70) showed strong selectivity for faces. Cells showed weaker but significant responses to apples, clocks, and other face-like objects, but not to any non-face-shaped object. Reverse correlation of spike trains from individual cells to a rapidly changing cartoon face parametrized by 20 facial dimensions showed that many neurons vary their firing along only one or two facial dimensions; important dimensions include pupil size, inter-eye distance, height of feature assembly, eyebrow slant, hair thickness, face shape, gaze direction, and face view. Degraded faces (low-contrast or partially occluded) elicited a longer latency and more sustained response compared to intact faces, suggesting a role for top-down and/or lateral feedback in determining whether an object is a face or not.

### **Mechanisms of Shape Processing in Macaque Area V4**

**Winrich Freiwald**

*University of Bremen and Harvard Medical School*

Shape processing along the ventral pathway critically depends on cortical area V4. Understanding shape processing in V4, therefore, has been an active area of research over the last twenty years. The goal of our study has been to identify some of the intermediate level processes and to relate them to underlying receptive field mechanisms. To address this question, we used single unit recordings in the fixating macaque monkey. Receptive field properties and shape selectivity have been studied with a set of experiments, all employing rapid stimulus presentation updates and reverse correlation analysis, but differing in stimulus content and complexity. In the first experiment, cells were stimulated with a sparse noise stimulus consisting of pairs of small squares of same or opposite contrast dots. These experiments revealed complex and often dynamic receptive field organization beyond what has been reported for early cortical areas. Shape selectivity was probed with experiments using gratings of Cartesian and non-Cartesian categories. All neurons exhibited marked within- and across-category tuning properties, with an overall population preference for circular shape - a preference which could be predicted from their second-order receptive field maps. Thus, we begin to understand shape selectivity in V4 neurons in terms of their receptive field mechanisms.