# MIT 2006 Progress Report Summary

## 1. Specific Aims

The MIT project continues to involve both computation and physiology in monkeys. Computational modeling of visual cortex interacts with experiments in the nearby physiology labs of E. Miller and J. DiCarlo (now one floor up). We continue to be guided by the model of recognition, which is itself evolving as an effect of the experiments, in our efforts to understand (a) the properties of selectivity and invariance of recognition, especially with image clutter, in IT and PFC cortex of behaving macaque monkeys and (b) the relations between identification and categorization. The model itself is improving tremendously every year thanks to the collaborative projects and interactions fostered by the Conte Center.

DiCarlo and Poggio are testing the effects of clutter on the selectivity and invariance of IT neurons and how to explain the results in terms of the model of object recognition. Miller and Poggio (with Riesenhuber) are investigating the neural bases of the recognition tasks of *identification and categorization.* In addition, Poggio (with Koch and Ferster) is working on biophysically plausible circuits for the two key operations in the recognition model – the max-like operation and the Gaussian-like multidimensional tuning. The work involves a close collaboration with CalTech on the computational side and with Northwestern on the experimental side.

Our specific aims are listed below from the original proposal:

**Aim A.1:** To determine the baseline IT neuronal relationships of a) shape-selectivity and clutter-tolerance, and b) position-tolerance and clutter-tolerance.

**Aim A.2:** To re-examine the relationship of shape-selectivity and clutter-tolerance and the relationship of position-tolerance and clutter-tolerance in the same monkeys after extensive training in clutter.

**Aim B.1:** To determine if there is a common neural substrate for different recognition tasks.

**Aim B.2:** To study the neural bases of the interaction of identification and categorization in Categorical Perception.

Since the beginning of our project, we have added four (three last year, one this year) new collaborative aims and deemphasized A.2 (see later):

**Aim N.1:** To explore the mechanisms underlying the max-like as well as the Gaussian-like tuning operations in cortical cells.

**Aim N.2:** To refine the model and check its prediction about properties of V4 cells (using data from Reynolds and Desimone and especially in ongoing collaborations with the Harvard lab of Dr. Livingstone and the JHU lab of Dr. Connor).

**Aim N.3:** To test the performance of the model on complex, natural images and compare it to the performance of humans and the response of neurons in the monkey IT and PFC.

**Aim N.4:** To predict and test the invariances of the representation at the level of neural activity in a population of IT neurons under presentation of specific objects.

## 2. Studies and Results

**"New" Aim N.1:** We planned to study computationally (with CalTech) and experimentally (with Northwestern) the circuitry underlying the max operation and tuning implemented as a normalized dot product. As a proof of concept, we have developed models of local circuits utilizing shunting inhibition for both operations in two versions, a) with non-spiking and b) with spiking neurons. These models compute a good approximation to either the normalization or the max operation for two inputs with very similar architectures, providing predictions that are testable experimentally. Continuing to refine these circuits and investigating the match between models and experiments, we are dealing with the question of neural coding, i.e. what are the constraints for these circuits in terms of time and size of the neural population for representations that are effectively between a firing rate and a temporal code.

**"New" Aim N.2:** We are investigating the feature selectivity within the intermediate visual areas before IT (such as V4 and V2), by replicating the tuning properties of the neurons to various visual stimuli. While incorporating more realistic normalization-based Gaussian-like tuning (Kouh and Poggio, 2004) in the model, we have achieved two major results:

a) We developed a systematic methodology to quantitatively fit the shape selectivity of single V4 neuron with a model unit (Cadieu *et al*., 2005). Using this technique, we have shown that we can successfully fit different, independent data sets on V4 from several physiological groups, i.e., the position-specific tuning for boundary conformation (Pasupathy & Connor, 2001), the tuning to gratings and sparse noise stimuli (Freiwald & Livingstone, 2005) as well as data on the weighed average effect in the absence of attention (Reynolds, Chelazzi & Desimone, 1999). With the currently available data, the estimation of model parameters is under-constrained and produces multiple solutions. However, the obtained set model units produces consistent shape tuning and invariance properties, and the cross-validation shows that their selectivity can *generalize to unseen stimuli within the stimulus set*. Furthermore, we can use the set of candidate model units to understand how the observed selectivity may arise from a feedforward mechanism and to make predictions to a new set of stimuli. We are currently preparing a manuscript of fitting boundary-conformation tuned neurons, in collaboration with A. Pasupathy at MIT and C. Connors at JHU.

b) One of our most significant results is that the unsupervised developmental-like learning stage (see New aim N.3) during which model units learn their selectivity from natural images produces a population of tuned units whose shape-tuning statistics are very similar to those of V4 neurons tested on several stimuli dataset.

Our work thus provides possible constraints and testable predictions on V4 tuning, driven by the demands of object recognition in natural scenes, and it also suggests an interesting alternative to descriptive models (e.g., Pasupathy & Connors, 2001) and low-level theories (e.g., Gallant *et al*., 1996) of area V4. We are continuing to collaborate with V4 physiologists (Connors, Freiwald, and Livingstone) on both the modeling of existing data and the design of new experiments to gain further understanding of the shape tuning properties in the intermediate areas of the ventral stream.

**"New" Aim N.3:** Over the last two years we extended significantly the model of the ventral stream in visual cortex. The model is now closer to the anatomy and the physiology of visual cortex with more layers (reflecting PIT as well as V4) and with a looser hierarchy (reflecting the bypass connections from V2 to PIT and V4 to AIT). It has also significantly more units, which was made possible by a new learning stage which provides a generic dictionary of shape-components from V2 to IT and thereby a rich representation that can be used by the task-specific categorization circuits in higher brain areas. This vocabulary of tuned units is learned from natural images during a developmental-like, unsupervised learning stage in which each unit in the S2, S2b and S3 layers becomes tuned to a different patch of a natural image. The resulting dictionary is generic and universal in the sense that it can support several different recognition tasks, particularly the recognition of many different object categories.

The new model also agrees with other data in V4 about the response of neurons to combinations of simple two-bar stimuli (within the receptive field of the S2 units). Some of the C2 units in the model show

a tuning for boundary conformations which is consistent with recordings from V4 (see New aim N.2). Read-out from C2b units in the model predicted (Serre *et al.*, AI MEMO 2005) recent read-out experiments in IT (Hung *et al.*, 2005), showing very similar selectivity and invariance for the same set of stimuli. Additionally the tuning of the S4 units is compatible with IT data in the presence of clutter (see Aim A.1).

We also found that learning improves drastically the recognition performance of the model in clutter (aim A.2). Not only can the model duplicate the tuning properties of neurons in various brain areas when probed with artificial stimuli, but it can also handle the recognition of objects in the real-world: the model performs much better than other less detailed biologically motivated feedforward models — and performs at least as well as state-of-the-art computer vision systems which do not have any relation with the anatomy and the physiology of the ventral stream (Serre *et al*., CVPR 2005, Serre *et al*., PAMI 2006).

We recently compared the performance of the model and the performance of human observers in a rapid animal vs. non-animal recognition task for which recognition is fast and cortical back-projections are likely to be inactive. Results indicate that the model predicts human performance quite well when the delay between the stimulus and the mask is about 50 ms. This suggests that cortical back-projections may not play a significant role when the time interval is in this range, and the model may therefore provide a satisfactory description of the feedforward path (Serre, Oliva & Poggio, "A feedforward theory of visual cortex predicts human performance in a rapid categorization task", in preparation).

**"New" Aim N.4:** A remarkable ability of our visual system is the possibility of recognizing objects under many different views and transformations. We planned to directly quantify the ability of populations of units from different stages of the model to decode information about complex objects and compared the results against recordings in IT using the same stimuli.

We recently used a biologically plausible statistical classifier to quantitatively characterize the information represented by a population of IT neurons about arbitrary complex objects in macaque monkeys [Hung *et al.*, 2005a]. We observed that we could accurately (performance > 90%) read out information about object category and identity in very brief time intervals (as small as 12.5 ms) using a small neuronal ensemble (approximately on the order of 100 neurons). The performance of the linear classifier that we used for decoding could, at least in principle, correspond to the information available for read-out by targets of IT, such as a neuron in PFC [Miller, 2000]. During such short 12.5 ms intervals, neurons typically conveyed only one or a few spikes, suggesting the possibility of a binary representation of object features. Importantly, the population response generalized across object positions and scales.

The observations obtained upon recording from populations of IT neurons are in agreement with the predictions made by the current theory and there is a quantitative agreement between the model and the observations obtained from the recordings in IT. In the near term we plan to use the model to quantitatively explore novel questions. In particular, we plan to use the model to explore several new scenarios including invariance to background changes for object recognition, invariance to presence of multiple objects and extrapolation to large numbers of objects and categories.

**Aim A.1:** We have made significant progress on this aim and we have recently published a study (Zoccolan *et al*., 2005) in which we examined IT responses to pairs and triplets of objects in three passively viewing monkeys that have been previously trained in object recognition tasks using isolated visual stimuli only. We have collected response data from IT neurons in all three of those monkeys (104 IT neurons overall). Probing neurons with objects spanning a broad range of effectiveness, we focused on the neurons whose responses were selectively tuned across different shape sets. One monkey was tested using sets of parameterized stimuli (cars, faces, and abstract silhouettes) with defined shape similarity, which were presented alone or in pairs in two different locations (1.25º above fixation and 1.25º below fixation). Two other monkeys were tested using a fixed set of simple geometrical stimuli (star, cross, and triangle) presented in all possible single, pair and triplet combinations in three different locations (2º above fixation, at fixation, and 2º below fixation). The shapes were scaled to fit within a 2º bounding circle.

Almost all IT neurons in our study showed response suppression when a second (clutter) object was presented along with a preferred object, no matter how dissimilar the clutter object was from the preferred object, while the original HMAX version of the model (Riesenhuber & Poggio, 1999) predicted recovery from suppression for "sufficiently dissimilar" clutter objects, (i.e. prediction of a U-shaped relationship between clutter tolerance and shape similarity of the clutter objects to the preferred object). We now realize that the U-shape prediction was a poor choice in the proposal, since it depended crucially on finding units which receive inputs from C2 cells activated strongly by one object (the preferred one) but not by the "sufficiently dissimilar" one. It turns out that such a situation is unlikely even in the old HMAX version of the model. The present version of the model -- which is more realistic in terms of units and areas (see earlier) -- suggests that it may be difficult to find neurons (and patterns) that show a U effect, because almost any object will activate to some degree a significant number of the inputs to IT (the hypothesis in the current version of the model is that there are hundreds to thousands of effectively "binary" afferents spanning a broad range of selectivity and invariance).

Our monkey recordings showed that, for the neurons *selected* under the conditions described above, a large fraction of each neuron's responses to multiple objects could be reliably predicted as the *average* of its responses to the constituent objects in isolation. In particular, the agreement of neuronal data to this "average effect" becomes extremely good when responses of even a small population of these selective IT neurons were pooled. These findings are potentially important because an exact average affect for all neurons and all patterns would strongly disagree with the predictions of the model (both the old and the new version). For this reason, following the completion of the above experiment, we started: 1) a new series of experiments aimed at probing the generality of the "average effect" over a "random" population of IT neuron and in parallel, and 2) a new series of simulations of the recognition model.

1) The new series of recordings is from a population of IT neurons spanning a broader range of shape selectivity, compared to the previous study. Instead of focusing on highly selective neurons tuned to small sets of similar shapes, as done in the previous study, we are currently measuring the selectivity of IT neuronal responses over a large set (~200) of natural objects. For each neuron, clutter tolerance, receptive field size, contrast sensitivity and tolerance to size changes are also measured. The first set of recordings (~60 neurons) has been completed from one monkey (already involved in the previous study). These new recordings show that most IT neurons do not show an average effect and some have responses that are more robust to clutter than expected from the "average effect". Interestingly, such deviations from the "average" seems to correlate well with the amount of selectivity of the recorded neurons, with an inverse relationship between clutter tolerance and shape selectivity, in good qualitative agreement with the new model simulations (see below).

2) The simulations have the goal to better understand the computational mechanisms that may produce the "average effect" observed in the data and to further explore which ranges of model parameters can affect the clutter tolerance of model neurons (the model suggests that averaging across units in IT should show an approximate average effect). These simulations revealed that the number of afferent units to which an IT model neuron is connected plays a key role in determining both the level of clutter tolerance and shape selectivity of the neuron. More specifically, increasing the number of afferents, the selectivity of an IT model neuron increases while its clutter tolerance decreases. In addition to the number of afferents, the amount of activation of such afferents to the presentation of the preferred object of the neuron affects the clutter tolerance, so that neurons connected to the afferents that are strongly activated by the preferred stimulus show higher clutter tolerance. Such simulation results suggest that there are different parameters and conditions that affect the selectivity and clutter tolerance of a neuron and that there may be a range of clutter tolerance levels.

Our new sets of experiments and simulations focus on the core of Aim A.1, since they are aimed at understanding the relationship between the object selectivity of an IT neuron and different kinds of tolerance to stimulus transformations, such as presence of clutter objects and changes of position, size or contrast of the "preferred" objects. The trade-off between selectivity and tolerance we are exploring is a key feature of the primate visual recognition system. The model has been instrumental in questioning the generality of the initial results.

**Aim A2:** Because we sought to first understand the 'baseline' clutter tolerance properties of IT neurons, we have not given our monkeys extensive training in the same recognition task in the presence of the distractor or other clutter objects. However, to lay the groundwork for understanding the effect of experience on IT clutter tolerance, we have: 1) completed the first characterization of the 'baseline' IT responses in clutter for neurons selectively tuned in small sets of similar shapes (animals not trained in clutter, described above); 2) started a new series of experiments to probe our findings over a population of IT neurons spanning a broader range of shape selectivity. We are also adapting parameters in the present version of the model using the new experimental findings on selectivity and clutter invariance.

**Aim B1:** To further explore whether there is a common neural substrate for different recognition tasks, we have expanded our comparison to other forms of categorization/recognition tasks. We used data from another project in which monkeys were trained on a version of our morph task that employed four prototypes and two orthogonal boundaries that carved up this space in two mutually exclusive category schemes. Of the third of responsive, randomly selected, PFC neurons (96/329, or 29%, at $P<0.01$) with category effects, about half of the neurons showed selectivity for one category scheme, and half for the other. This was in sharp contrast with the categorization/recognition task reported in last year's progress report in which monkeys seemed to adopt a "hybrid" strategy in which the monkeys simultaneously encode the category membership as well as the identity of individuals. This may be because there was a high degree of interference between the different category schemes and thus, the PFC orthogonalized the representations to minimize error. We are also using a classifier analyses to determine if we can detect any subtle differences in neural representation in the categorization versus recognition task.

**Aim B2:** The state of this project did not change. Monkeys are currently being prepared for this aim (Roy, Miller). This task requires that monkeys be over trained on the categorization task in order to produce effects of categorical perception. So, the monkeys used for Aim B1 will be used for this aim once the experiments for Aim B1 are completed.

## 3. Plans

**Aim 1:** We have recently published a study (Zoccolan *et al*, 2005) with a first assessment of the clutter tolerance properties of IT neurons and we are currently extending this characterization to populations of IT neurons with a broader range of shape selectivity and using a large set of natural objects. At the same time, we are simulating the behavior of IT model neurons under analogous clutter conditions, using the same set of stimuli. We completed our new set of experiments from one monkey subject and we recently started to record from the second monkey. We expect these recording to be completed within the next few months. These data will provide new insights about: 1) the trade-off between object selectivity, clutter tolerance and position tolerance in IT; 2) the impact of clutter objects on IT neuronal responses as function of their distances (in the visual field) from the preferred object and their contrast; 3) a population measure of contrast sensitivity in IT.

We will also perform new simulations using the object recognition model to better understand which computational mechanisms are essential to produce the inverse relationship between clutter tolerance and object selectivity, which we observed both in the recorded and simulated IT neuronal responses. In particular, we will focus on the key computations of the model (maximum-like invariance and Gaussian-like tuning operations) to test if both of them are essential and if some of them can be replaced by alternative and, possibly, simpler computations. In addition to matching the qualitative behavior with the IT neurons (i.e., tradeoff in selectivity and invariance), we will use the data to quantitatively constrained model parameters at the level of IT. Furthermore, we will study the possible computational function of the relationship between clutter tolerance and object selectivity.

We expect to have a paper that will be ready for submission within the next several months and that will contain both the experimental and theoretical results outlined above.

**Aim A.2:** Because of the unexpected observations described in Aim A1 (above), we have focused our efforts on understanding IT response properties in clutter without extensive training (Aim A1).

# Caltech 2006 Progress Report Summary

## 1. Specific Aims

The Caltech project is organized around the central theme of attentional aspects of object recognition, using visual psychophysics and computational and biophysical modeling. The research is organized into two aims: (1) Psychophysics of attention and recognition in natural scenes to dissect bottom-up from top-down components. This will determine the limits of the current feed-forward recognition and saliency models. (2) Integrate our saliency model with the feed-forward recognition system central to our Conte Center to implement attentional modulation of object recognition.

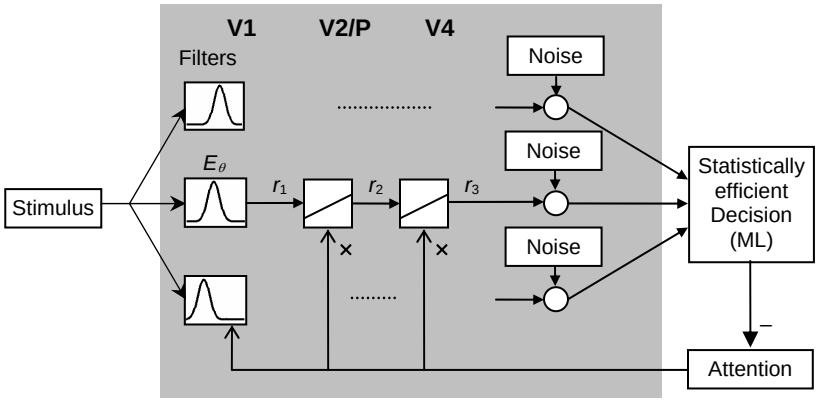
## 2. Studies and Results

**Top-Down attention enhances sensory processing in human extra-striate visual cortex** (Farshad Moradi & Christof Koch):

Selective attention evokes fMRI response in visual areas as early as LGN and V1 even in the absence of retinal input, but it is unclear whether stimulus-dependent activation also increases with attention. Here, we tried to dissociate stimulus-dependent, attentional effects from changes in baseline BOLD response by directing subjects focal attention either before, or after, a peripheral grating was displayed. Two square-wave gratings were displayed in upper screen quadrants at 10 degree eccentricity. A cue instructed subjects to attend to either one of the two gratings, or---more frequently---to a foveal attention demanding task. The cue appeared either 400 ms before, or 250 ms after, the onset of the stimuli. If the target was one of the gratings, observers reported whether it is tilted to the right, or to the left. By varying the orientation of the grating we examined whether or not pre-cuing improves discrimination compared to post-cuing.

Single neuron recordings in monkeys and human visual event related potentials indicate that facilitation of neural responses to stimulation mainly occurs in extra-striate cortices, and not much in V1. Consistent with those results, we demonstrate that if attention is deployed before the peripheral target is displayed, V4 response and subjective discrimination are both enhanced compared to when the targets appears first. Contrariwise, attending before the target appears does not enhance V1 activity.

These findings are in agreement with the hypothesis that attention facilitates subsequent sensory processing by increasing neural gain in extrastriate cortex in a top-down manner. A simple gain cascade model that assumes selective attention increases both the gain and the baseline BOLD activity in V1, V2/VP and V4 was used to explain the results (Fig 1). Our suggested model has four features: (1) In the absence of attention, gain < 1. Therefore the response to the peripheral pattern decays as it travels from V1 to V4. (2) Pre-cueing increases the gain, thus the response increases from V1 to V4. (3) In the post-cue condition, gain increases after the initial feed-forward activity has subsided. Therefore the increase in the baseline signal dominates the BOLD response. (4) Discriminability depends on the onset ($t_0$), and duration ($t_1$–$t_0$) of the attentional window with respect to the time constant of the decay of activity ($\tau$).




**A visual search task overrides static – but not dynamic - bottom-up biases to human overt attention** (Wolfgang Einhäuser & Christof Koch):

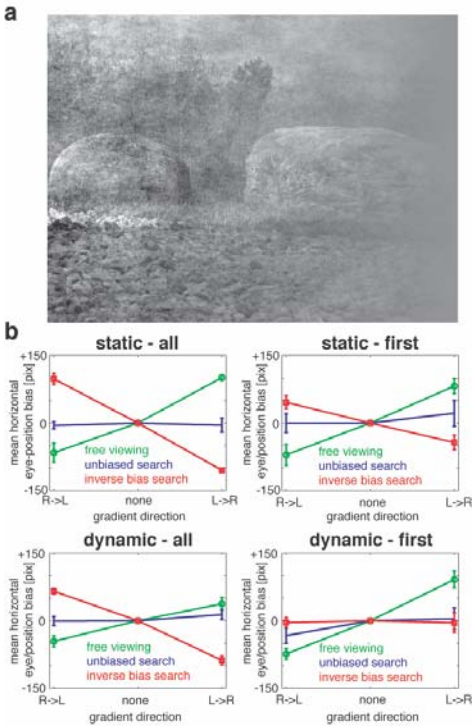
While quantitative approaches to attention are successful in predicting fixation patterns, their "bottom-up" description remains incomplete: attention is to a large extent guided by other ("top-down") factors, such as the task. We investigate how the task directly modulates bottom-up biases in selective, focal attention.

We measured eye-movements with a non-invasive infrared eye-tracker (EL1000, SR Research), while they viewed grey-scale, natural outdoor scenes. Different levels of noise were added to the Fourier phase of the stimulus, making it look less "natural", while preserving its amplitude spectrum. We imposed two different modifications, either smoothly modulating contrast from one side of the image to the other ("static") or alternating one side at 5Hz between up- and down-modulated contrasts ("dynamic").

In both conditions, observers performed two different tasks: in "free-viewing" (block 1 and 4), observers had to judge whether the image appeared "natural" or not; in "search", observers had to answer the same question but additionally, had to find and fixate a small "bull's eye" target somewhere in the image as quickly as possible. Fixating the target terminated the trial immediately. In an "unbiased search" condition (block 2) the target occurred with equal probability on either side of the stimulus. In an "inverted bias search" condition (block 3) – unbeknownst to the observer – the target occurred always on the side of lower contrast (static) or the non-flickering side (dynamic).

In free viewing, mean fixation location is robustly biased toward the side of higher contrast or of flicker – to a similar extent in both conditions. In the search task, these biases are absent (unbiased search) or even reversed (inverted bias search). In the static condition, this top-down "overriding" of the bottom-up bias is present at the first fixation after stimulus onset, while in the dynamic condition, the bias induced by flicker is still partly present at this early stage. We conclude that bottom-up factors controlling subject's gaze, for instance via a saliency-map, can be rapidly overridden by top-down, task-demanding factors.

The figure shows **(a)** an image with a medium-noise level added and with a right-to-left static bias (contrast gradient) superimposed. The central cue and the target (right on top of the left stone) are visible **(b)** Bias of horizontal fixation. Taking all fixations into account (left column), the bottom-up bias present in free viewing (green, mean and s.e.m. over 5 subjects) is overridden by a spatially unbiased visual search task (blue) and reversed if the target consistently appears in the less salient region of the stimulus (red). Taking only the first fixation into account (right column), the overriding is already present and the inversion visible for a static bottom-up bias (contrast increase, top row). For a dynamic bias (5Hz flicker, bottom-row), the overriding is not fully deployed and the inversion is not observable until later.

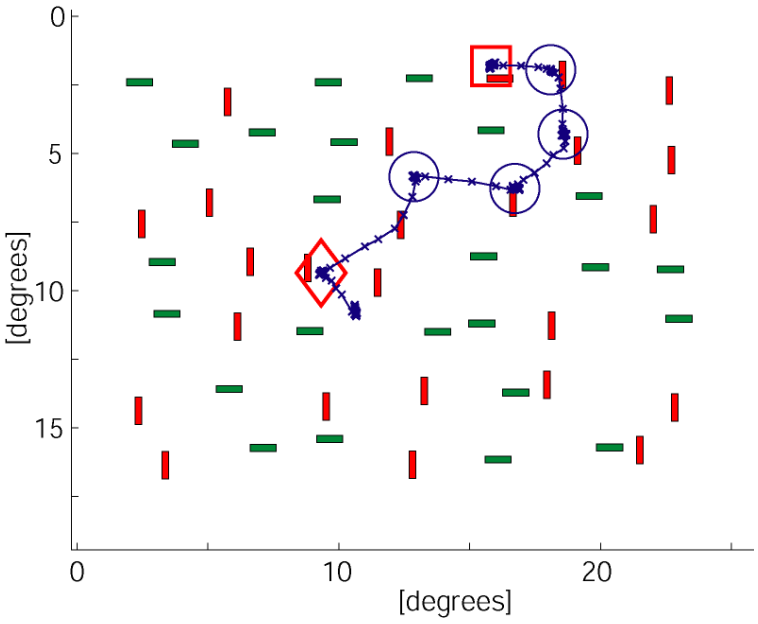



**Deployment of feature based top-down attention during visual search** (Ueli Rutishauser & Christof Koch):

Task-specific information can strongly influence the response of neurons in the visual system. It is, however, unclear which information about a task is used to bias neuronal responses in a top-down manner. To investigate this in a quantitative manner, we developed a generative model that reproduces eye movements that we measured during the performance of simple visual search tasks. We calculated the conditional probabilities that observers fixate, given the target, near an item in the display sharing a specific feature with the target instruction. We use the conditional probabilities as a measurement of which features were biased by top-down attention. Additionally, we use the number of fixations it took to find the target as a measurement of difficulty. From this, we infer the strength of top-down attentional modulation of 3 features: color, orientation and size. Further, we find that only a model that biases multiple feature dimensions in a hierarchical manner can account for the data. Biasing multiple features implies that the summation of different feature dimensions requires a sum-type rather than max type operation. These findings are very compatible with neurophysiological findings of the firing properties of neurons in V4 and the frontal eye fields (FEF). The model can be used to predict the extent of gain modulation of neurons in V4 and FEF from eye tracking data.

The figure shows one trial of a search for a red, horizontal target with the fixations (circles) and eye movements made by the subject. The first fixation is indicated by the red diamond, the last (on the target) by the red rectangle. Notice that all the distractors that are fixated by the subject share the color but not the orientation with the target.

**Error!**




### 3. Plans

Current evidence from our and other laboratories suggests that purely bottom-up, saliency-driven models of gaze allocation can explain a significant fraction of scan-path data in both artificial and natural images. However, it appears that when subjects have to search for a particular item, a purely saliency-driven approach can be overridden in favor of a much for guided search. The challenge for the coming year will be to construct models of attention, for instance by modulating a central saliency-map by feedback connections from frontal-eye-fields or other regions in prefrontal cortex, to better account for gaze control in natural images using both bottom-up as well as top-down features.

# Georgetown 2006 Progress Report Summary

## 1. Specific Aims

The Georgetown subcontract was originally part of the MIT project, the subcontract arising from Dr. Riesenhuber's move to Georgetown University to start a lab there. Thus, activities at Georgetown are directly related to the aims of the original MIT project, *i.e.,* its **Aim A** (investigating the neural mechanisms underlying object recognition in clutter) and **Aim B** (to determine if there is a common neural substrate for different recognition tasks). In particular, work in the Georgetown project has focused on continuing our effort to apply the computational model at the core of our center to quantitatively model not just physiological data but also human object recognition performance and brain activity as measured by fMRI.

Based on our findings (see below) that have shown the success of this approach, we have added a new goal (**Aim NG:** To study the neural bases of single-word reading as an object recognition problem, using an integrative approach of computational modeling, behavioral techniques, and fMRI) that now breaks new ground by investigating how much of single-word reading, a uniquely human skill whose mechanistic neural bases are still poorly understood, can be accounted for by our computational model of object recognition. The acquisition of reading skills strongly relies on high-level neural plasticity, and reading itself involves bottom-up as well as top-down processes, providing a novel domain, highly relevant for clinical applications, in which to study the basic science questions central to our Conte Feasibility Center.

## 2. Studies and Results

**Aim A:** As described in last year's progress report, to test model predictions regarding object recognition in clutter in humans using behavior and fMRI, we first extended the model to model fMRI and behavioral data. Based on published physiological data on monkey face cell tuning specificity, brain imaging data on FFA (fusiform face area) selectivity from fMRI, and human behavioral data on face discrimination performance, we have developed a computational model of face neurons in the FFA. The simulations showed that the data on face processing can be well accounted for in our standard modeling framework as a result of extensive experience with faces, without having to postulate additional "face-specific" mechanisms, in line with our earlier psychophysical results. We then used the model to quantitatively and predictively link model face neuron responses, BOLD contrast in fMRI, and behavior. A key element of our experimental approach is that we are using fMRI rapid adaptation (fMRI-RA) paradigms to more directly test our hypotheses about neural tuning than possible with conventional techniques that look at the average BOLD-contrast response to individual stimuli. Our data show that we can indeed use fMRI-RA to study neuronal tuning with fMRI, and moreover that face discrimination appears to rely on neuronal mechanisms that are not different from those for other objects (see Jiang et al., *Neuron*, 2006). We have now started to investigate the original model hypotheses regarding the neural mechanisms underlying recognition in clutter using in fMRI (see "Plans" below).

**Aim B:** We have continued our collaboration with the lab of Earl Miller, investigating the neural bases of different recognition tasks (in particular, different categorization schemes) in monkeys. Details are described in the MIT Progress Report.

In parallel, we have developed a human subject version of the original categorization task used in the monkey studies, using morphed cars instead of the cat/dog morphs employed in the monkey study, to avoid confounds with subjects' preexisting categories for the animal stimuli. Using a morph space spanned by four prototypes, we have now trained 19 subjects on a categorization task. We recorded brain activation before and after training using fMRI-RA techniques, allowing us to study learning effects in the same subjects, over the whole extent of the brain. We find that training humans on the perceptual categorization task leads to the sharpening of a stimulus representation coding for the physical appearance of stimuli in lateral occipital cortex (LO), part of the lateral occipital complex, LOC, the human

homologue of monkey area IT, which has been postulated to play a key role in human object recognition. When subjects were judging the category membership of car images, bilateral prefrontal and parietal cortices were more strongly activated when the two cars belonged to different categories than when they belonged to the same category. Neurons in the right PFC (rPFC) exhibited the strongest category-selective activation and were sensitive to an explicit change of category membership and not merely to shape differences. This category-selectivity was greatly diminished to the extent that it could not be detected when subjects were doing a position displacement task. These results support the model and suggest that similar principles underlie category learning in humans and monkeys.

**Aim NG:** One research focus of our Center is to understand the interaction of bottom-up, stimulus-driven and top-down, task-specific effects in object recognition. In a new exciting research direction, we are now studying top-down influences in more detail in the very interesting domain of single word reading. Preliminary data, generated by a new graduate student in the lab, Laurie Glezer, suggest that word characteristics (e.g., whether a shown letter string is a real word or a pseudoword) can modulate activation in word processing areas, but that this modulation can be controlled for by masked presentation that prevents a rapidly presented word from reaching consciousness, while still activating reading-related areas. We have performed initial behavioral and fMRI experiments that suggest that word representation in the so-called "Visual Word Form Area" (VWFA) in left fusiform cortex follows similar principles as the representation of faces in the FFA or cars in LOC (for our trained observers), and we are currently studying the representational mechanisms in the VWFA.

## 3. Plans

**Aim A:** We will continue the psychophysical and fMRI testing of the model predictions for recognition in clutter, in particular the U-shape prediction for recognition performance (see MIT Progress Report), for which we have some intriguing but very preliminary initial evidence. In fact, measuring subject performance in clutter and doing whole-brain fMRI offer certain advantages over the measuring of single neuron response properties whose relevance for task performance is unclear. The human data should thus complement the monkey data very well.

**Aim B:** We will continue the collaboration with Miller's group, and also continue the human studies on categorization. In particular, we are interested in task-specific modulations and transfer of learning. Indeed our data suggest that categorization training also improved subjects' performance on car *discrimination,* also within the categories, supporting the split predicted by the model into a shape-specific but task-agnostic representation in IT/LOC that can be utilized for different recognition/categorization tasks involving the same stimuli. We will further investigate the dynamical development of category learning by imaging subjects also at intermediate stages of learning.

**Aim NG:** We will continue our investigation of top-down modulations of VWFA activation by looking at responses to unmasked and masked (i.e., not consciously perceived) words. We will further study tuning specificity in the VWFA to understand how task demands shape high-level plasticity in this area. In particular, we will vary the physical (orthographic) similarity and lexicality of the prime in relation to a high frequency target to examine neuronal tuning in the VWFA with an fMRI-RA paradigm. This will allow us to start building a model of VWFA tuning to serve as a starting point to generate hypotheses for the neural bases of reading deficits, similar to the face perception work described above.

# Northwestern 2006 Progress Report Summary

## 1. Specific Aims

In the previous period, we described our studies of spatial summation in the receptive fields of complex cells in area 17 of cat visual cortex. As discussed in the last progress report, our early data (Lampl *et al.*, 2004, *J. Neurophysiology*. 92:2704-13) were somewhat at odds with one of the classical papers on complex cells (Movshon *et al.*, 1978, *J. Physiol.*, 283:79-99). In the classical theory of complex cells, as first proposed by Hubel and Wiesel in 1962, spatial interactions between different parts of complex receptive fields of complex cells should resemble the receptive fields of simple cells, which are thought to supply the predominant excitatory synaptic input to complex cells. This model predicts that when bars of the same polarity (bright/bright or dark/dark) flashed close together in a complex cell's receptive field, they facilitated one another; bars presented farther apart should antagonize one another. The opposite should occur when the bars of opposite polarity (bright/dark) are used. These interactions, much like the interactions in the receptive fields of simple cells, occur even though the complex cell receptive fields are uniform in their responses to single bar stimuli. Movshon *et al*. confirmed this model in their small sample of complex cells.

Lampl *et al*. had found -- in contradiction to the Movshon work and the classical model -- that the interaction between bars in a flashed pair was independent of polarity or separation. The response was instead MAX-like: the response to pairs of flashed bars was similar to largest of the responses to the individual bars, no matter what the relative distance between the bars of the pair. In the progress report of the last period, we had extended the Lampl *et al.* result by finding that there were two seemingly distinct types of complex cells: those that behaved as Hubel and Wiesel's model predicted (as Movshon *et al*. had found), and those that acted in a more MAX-like fashion (as Lampl *et al.* had found). These two cell types emerged when we made more exhaustive tests of the spatial summation in complex cells, presenting all possible combinations of bar pairs across the receptive field. Examples of the two response types are shown in Figure 1.

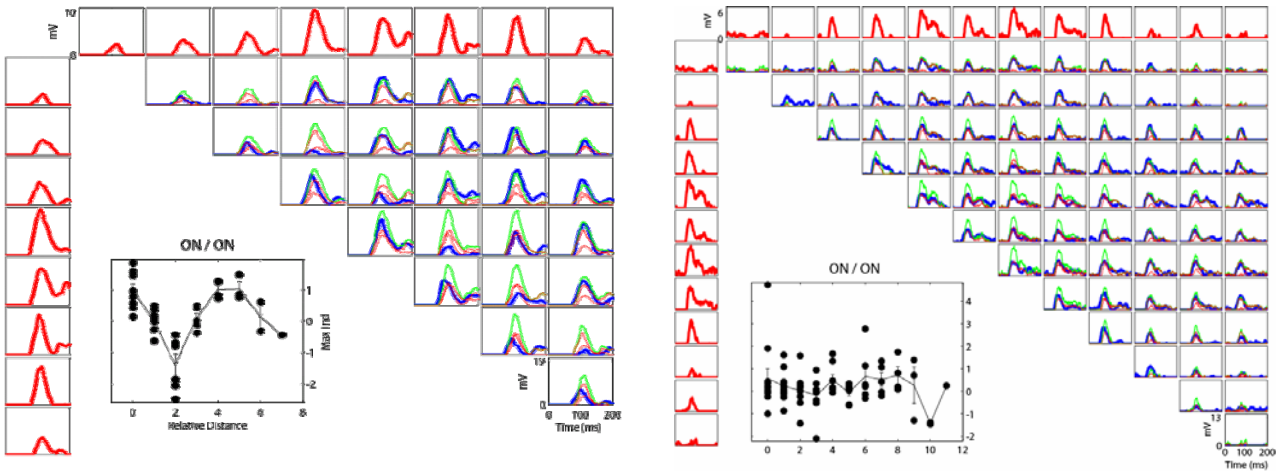



*Figure 1. Intracellularly recorded responses of 2 complex cells in area V1 of the cat visual cortex, a classic complex cell that conforms to the Hubel-Weisel model on the left, and MAX-like cell on the right. Top row and left column: responses to a briefly flashed, optimally oriented bar at 12 different positions within the receptive field. The rest of matrix shows the responses to simultaneously flashed pairs of bars (blue). Green traces show the sum of the responses to the two bars of the pair when presented individually. The inset shows the MAX index as a function of the relative distance between the bars in a pair for all pairs.*

## 2. Studies and Results

In the current period, we have extended these results by increasing our sample size and characterizing the sample quantitatively. For each cell recorded, we constructed an index of the MAX-like behavior of the cell from the inset graphs of Figure 1. The inset graphs show the MAX index of the individual responses to bar pairs as a function of separation of the individual bars in the pair. The maximum and minimum values of this index were extracted from the graphs and subtracted. The resulting measure of how much the MAX index varies as a function of bar separation was taken as an indication of how MAX-like the cell was. The more variability, the more the cell conformed to the Hubel-Wiesel model; the less the variability, the more the cell was MAX-like. This index was consistent within cells no matter whether bars of the same polarity or different polarity were used in the calculation. A histogram of this spatial variability index (SVI) is shown in Figure 2.

The histogram is significantly bimodal, suggesting that there exist two distinct types of complex cells. This distinction, it should be noted, is based on the synaptic input to complex cells (membrane potential measurements), not on measurements of their spike output (spike rate measurements). In contrast, we have recently shown that the distinction between simple and complex cells is bimodal only for the spike output of cortical cells, whereas the synaptic inputs form more a continuum along simple/complex spectrum. Spike threshold is therefore critical in establishing the simple/complex distinction. And yet these newly described types of complex cell seem to be established right at time of the synaptic input to the cells, and do not require threshold to be elaborated in the spike output.

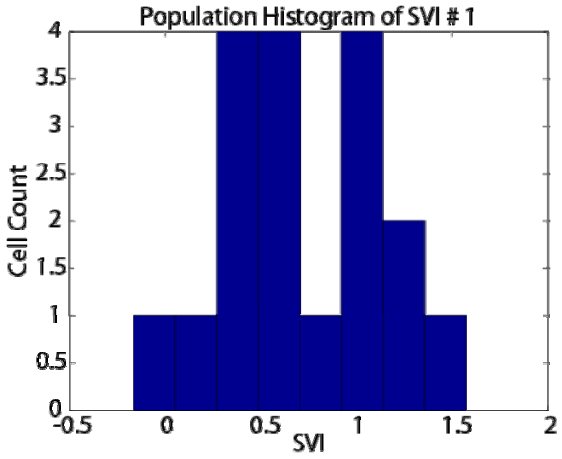



*Figure 2. A histogram of the spatial variation of the MAX index within complex cells. The histogram is significantly bimodal.*

In the previous period, we had seen that cells that are more or less MAX-like had different spatial frequency tuning. This finding has continued to prove true as we increase the size of our sample of intracellularly recorded cells. In Figure 3 width of spatial frequency tuning is plotted against the spatial variation index (MAX-like behavior). A significant correlation between these two measures is present.

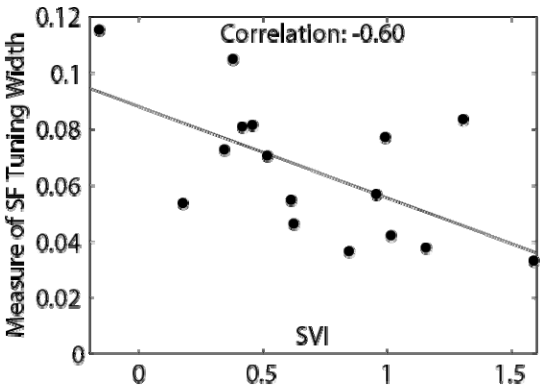



*Figure 2. The width of spatial frequency tuning for each recorded cell as a function of the spatial variability index of MAX-like behavior.*

We are currently testing whether the MAX-like and more classical complex cells differ from one another in other aspects of their responses. The latency of response to the flashed stimuli or the amplitude of responses, for example, seems not to differ significantly. Nor does receptive field size differ. We are also interested in length summation in the two types of cells. A previously described distinct population of complex cells is the so called special complex cells of Palmer and Rosenquist (1975, *Brain Res*., 15:27-42) and of Gilbert (1977, *J. Physiology.,* 268:391-421), which are known to project to the superior colliculus, have high spontaneous activity, and show length summation that saturates at a length far shorter than the length of the receptive field. We examining length summation, we hope to show whether the MAX-like cells correspond to the special complex cells or not. Finally, we are studying the laminar location of recorded cells and whether they receive direct, monosynaptic input from the lateral geniculate nucleus.

## 3. Significance

Complex cells are very difficult to study and characterize because they do not generally receive direct input from the lateral geniculate nucleus, and because their visual responses are highly nonlinear. Complex cells, however, constitute the bulk of the output of the primary visual cortex, generating the signals that V1 sends on to higher levels of cortex and forming the basis for higher levels of perception and object recognition. If our results continue to show the trends that we have identified so far, we will perhaps have described one of the first new distinct class of cells to be identified in some time. The MAX measure and the MAX model have given us an entrée into understanding complex cells at a new level of detail.

## 4. Plans

The next major task is to determine the synaptic and cellular mechanisms that distinguish MAX-like and classical complex cells. How is their different behavior generated by the cortical circuit? By recording the visually evoked changes in membrane potential while polarizing the cell with different levels of injected current, we can estimate the excitatory and inhibitory synaptic conductances underlying the potentials. These conductances will likely differ qualitatively in the two types of cells. Our colleague Ilan Lampl at the Wiezmann Institute, for example, suggests that MAX-like behavior might be produced by nonlinear summation of inhibitory inputs combined with linear summation of excitatory inputs. Other models of the mechanisms underlying the MAX-like behavior of complex cells are being developed by members of Dr. Poggio's and Dr. Koch's groups. Both the intrinsic properties of cortical neurons (synaptic and voltage gated currents) and the properties of the local circuit are being considered.

In addition we are currently working on increasing our technical capabilities for studying the synaptic input to complex cells. We have just completed construction of a 2-photon laser scanning microscope for *in vivo* recording of synaptically-triggered $Ca^{++}$ transients in cortical cells. We hope to be able to record the synaptic events that are evoked in different parts of a complex cell's dendritic tree. We will study how these signals vary across the dendritic tree, comparing cells that are identified in electrical recordings to be more or less MAX-like in their behavior.