Abstracts of NIH-Conte Meeting – September 11, 2006

Spiking Neural Circuits for Gaussian Tuning and the Maximum Operation

Jake Bouvrie and Ulf Knoblich
Massachusetts Institute of Technology

In an effort to realize an additional layer of biophysical plausibility in the model, we present preliminary spiking neural circuits which separately implement the maximum operation and an approximation to a normalized Gaussian tuning function. We investigate these circuits on two levels of complexity. First we simulate biophysically detailed networks of model neurons with realistic synapses and different network topologies in order to study their responses to different types of inputs and compare them to data from electrophysiological experiments. We then implement a simplified version of these circuits retaining the basic necessary mechanisms but decreasing their complexity in order to make them computationally less expensive. Our MAX design is based on a winner-take-all computation with integrate-and-fire elements, and exploits a tuned balance between self-excitation and all-to-all inhibitory dynamics. Gaussian tuning is approximated by delivering both divisive, normalizing inhibition and weighted excitatory signals to units operating in nonlinear regions of the spike-rate transfer function. In both cases, timing constraints are overcome by considering multiple parallel interacting copies of circuit inputs, outputs and networks themselves. It is our goal to eventually embed these circuits in the existing model and investigate the resulting firing patterns arising from the complex interaction between many simple subsystems.

Object Recognition in Clutter: Selectivity and Invariance Properties in the Monkey Inferotemporal Cortex

Davide Zoccolan and James DiCarlo
Massachusetts Institute of Technology

Visual object recognition has two core requirements: discrimination of each object from all other objects (shape selectivity), and generalizing over wide variation in each object's appearance (tolerance). Individual neurons at the highest primate visual processing stage (inferior temporal cortex, IT) are thought to have somehow attained both selectivity and tolerance. However, while either requirement can easily be obtained at the expense of the other, it is computationally challenging to achieve both at once. Thus, there is likely a tradeoff between selectivity and tolerance thorough the visual system, possibly including IT. On the other hand, it is possible that no such tradeoff exists in IT – for example, if IT neurons were tuned to highly specific "diagnostic" image features that might guarantee both high selectivity and high tolerance. Here we test these hypotheses by examining the relationship between selectivity and tolerance across a population of IT neurons. We found that IT neurons are not a homogeneous population of highly specific feature detectors, but instead span a broad range of shape selectivity. Most importantly, across this broad range of shape selectivity, we found that highly selective neurons are much less tolerant to changes in object appearance, including the presence of visual clutter, image contrast, and object position. Overall, these findings provide the first direct assessment of a core computational constraint that is likely to operate thorough the visual system (and possibly other sensory areas) forcing neurons to trade selectivity to gain higher tolerance.

Tradeoff Between Selectivity and Tolerance in Model of Object Recognition

Minjoon Kouh
Massachusetts Institute of Technology

One of the key principles in the model of object recognition is the leveraging of the tradeoff between selectivity and invariance, so that the complexity of tuning increases gradually and progressively along the feedforward hierarchy. In the model, the range of and the tradeoff between selectivity and clutter tolerance, which have been observed in physiology, are found in a variety of conditions. For example, the sharpness of tuning, the number of afferent neurons, the selection rule of the afferents, and other factors affect the tuning behavior.

Explaining Rapid Categorization

Thomas Serre, Aude Oliva and Tomaso Poggio
Massachusetts Institute of Technology

Here we show that the model can predict the pattern of performance achieved by human observers on a rapid animal vs. non-animal categorization task. We generated a balanced set of stimuli by selecting animal images from four different subcategories based on body size and viewing distance from the camera (from heads to full bodies in clutter). Recognition performance by human observers (n = 21) was tested with a backward-masking paradigm, i.e., 20 ms stimulus presentation followed by a variable inter-stimulus interval (ISI) then a 80 ms mask duration.

We found that the feedforward model could predict the pattern of performance of human observers (both HIT and d') on the different animal subcategories for an ISI of 30 ms (overall correlation between model and human observers $\rho = 0.72$, $p < 0.01$). To further challenge the model we tested the effect of image rotation on recognition performance. Consistent with previous psychophysics results (Guyonneau *et al*., ECVP 2005), both human observers and the model were fairly robust to image orientation.

On the Limits of Feed-forward Processing of Visual Information

Gabriel Kreiman
Massachusetts Institute of Technology

Many visual recognition tasks can be accomplished quite rapidly. This has inspired many of us to suggest that, as to be expected, back-projections are required during several recognition tasks in real-world situations. In line with this notion, a model of the ventral visual stream consisting entirely of feed-forward projections can perform at human level in rapid categorization tasks and it can mimic the initial responses in IT cortex during the recognition of single isolated objects. By combining the modeling efforts with a statistical classifier to read out from populations of neurons or model units, I will describe our explorations of more complex recognition scenarios involving objects embedded in complex backgrounds and multiple objects per image. These scenarios illustrate some of the limits of feed-forward visual processing and suggest that back-projections may be generally required during recognition in real-world situations.

Measuring the Cost of Deploying Top-down Attention

Dirk Walther (with Fei-Fei Li)
California Institute of Technology and York University

In many everyday situations we bias our perception from the top down, based on a task or an agenda. Frequently, this entails shifting attention to a specific attribute of a particular object or scene. To explore the cost of shifting top-down attention to a different stimulus attribute, we adopt the task switching paradigm, in which switch and repeat trials in mixed task blocks are contrasted with single task blocks. Using two tasks that relate to the content of a natural scene in a gray-level photograph and two tasks that relate to the color of the frame around the image, we were able to distinguish switch costs with and without shifts of attention. We found a significant cost in reaction time of 20-28 ms for task switches that require shifting attention to other stimulus attributes, but no significant switch cost for switching the task within an attribute.

In a feed-forward, hierarchical model of the ventral pathway, switching top-down attention to a different feature value within the same stimulus attribute requires biasing feed-forward connections at fairly high levels, e.g. in inferotemporal cortex (IT) or even the connections from IT to prefrontal cortex (PFC) for object categories. Thus, the effect of top-down attention to object categories could be interpreted as switching one set of synaptic weights for another one.

In contrast, when switching to a different stimulus attribute, processed by a different visual area, one would assume that task-specific biasing of neural activity has to happen at an earlier stage in the hierarchy, before specialization of processing streams takes place. In the case of switching between color and object detection, this could be area V4, V2 or even V1. Our current results indicate a higher cost in RT for task switches between attributes, i.e. for biasing at an earlier stage, than within attributes, i.e. biasing at a later, more specialized stage. This finding agrees with ideas of a reverse hierarchy.

Task-Demands Can Reverse the Effects of Visual Saliency

Wolfgang Einhäuser and Christof Koch
California Institute of Technology

While stimulus features and the task-demands affect observers' gaze allocation, the exact relationship between these "bottom-up" and "top-down" factors remains controversial. We recorded eye-movements in observers viewing photographs of outdoor scenes. Noise was added to the Fourier phase of images and observers always reported whether the stimulus appeared "natural". To bias bottom-up saliency, image contrast was increased along a horizontal gradient running smoothly from one side to the other. When observers only performed the natural/non-natural judgment, their eye-position was immediately biased to the high contrast side. This bottom-up bias vanished entirely, when observers had to detect a target hidden with equal probability in either side of the image. When the target – unbeknownst to the observer - always occurred on the low-contrast, i.e. low-saliency, side, all fixations were biased to this side, entirely reversing the bottom-up effect. In conclusion, while bottom-up saliency determines to which side subjects look under free viewing conditions, this bias can be overridden or even countermanded by top-down instructions.

Deployment of Feature-based Top-down Attention during Visual Search

Ueli Rutishauser
California Institute of Technology

Where the eyes fixate during search is not random; rather, gaze reflects an expectation of where the subject expects the target to be. It is not clear, however, what information about a target is used to bias the underlying neuronal responses. We engaged subjects in a variety of simple visual search tasks while tracking their eye moments. We derive a generative model that reproduces these eye movements and calculate the conditional probabilities that observers fixate, given the target, on or near an item in the display sharing a specific feature with the target. We use these probabilities to infer which features were biased by top-down attention: color seems to be the dominant stimulus dimension guiding search, followed by object size and, lastly, orientation. We use the number of fixations it took to find the target as a measure of task difficulty. We find that only a model that biases multiple feature dimensions in a hierarchical manner can account for the data. Contrary to common assumptions, memory plays almost no role in search performance. Our model can be fit to average data of multiple subjects or to individual subjects. Small variations of a few key parameters account well for the inter-subject differences. The model is compatible with neurophysiological findings of V4 and FEP neurons and predicts the gain modulation of these cells.

Differential Attentional Effects in Human Visual Cortex Consistent with a Feed-forward Gain Cascade Model

Farshad Moradi and Christof Koch
California Institute of Technology

Attention evokes fMRI response in striate and extrastriate areas even in the absence of retinal input, but it is unclear whether it enhances stimulus-dependent activation. We tried to dissociate stimulus-dependent effects from changes in baseline BOLD response by cueing attention either before, or after, a peripheral grating is displayed. Attentional effects that are independent of the input might be observed in both conditions whereas the enhancement of feedforward activity requires that attention is cued prior to the stimulus onset. Both cue conditions increased BOLD activity in V1-V4 compare to when attention was distracted. V1 activity was stronger when the cue followed the target compared to when the cue was presented first. However, the stimulus-dependent effects increased in subsequent visual areas and V4 activity was stronger in the pre-cue condition. A simple cascade model that assumes selective attention increases both the gain and the baseline activity in V1, V2/VP and V4 quantitatively reproduces the results. Such a model is compatible with human EEG and monkey single cell studies, reconciles them with fMRI studies, and explains the improved discrimination performance with attention. Measuring the Cost of Deploying Top-down Attention.

Reading Out Information From Populations of ITC and PFC Neurons

Ethan Meyers
Massachusetts Institute of Technology

Most current analyses of neural data focus on the properties of individual neurons and do not examine how information is coded in a more distributed manner through a population of neurons in a given brain region. By extending the neural readout techniques (Hung, Kreiman, *et al.* 2005) we are currently reanalyzing data that was recorded from ITC and PFC while a monkey was engaged in a delayed match to sample task (Freedman *et al.* 2003). Preliminary results give insight into the different information contained in ITC vs. PFC, into the information contained in individual neurons vs. a larger population of neurons, and into how the information content changes over the course of a given trial. In particular, PFC contains little information about the individual identity of a particular object. Also, the population of ITC neurons contains more information relevant to categorization while a given stimulus is visible, than any of the individual PFC neurons does at any time period.

Categorization Training Results in Shape- and Category-Selective Human Neural Plasticity

Xiong Jiang and Maximilian Riesenhuber
Georgetown University

Learning to categorize objects is a fundamental cognitive ability and an important domain in which to study the general neural mechanisms underlying cortical plasticity, as it requires combining “bottom-up” stimulus-driven information with “top-down” task-specific information. Findings from computational, behavioral and monkey single-cell recording studies suggest a two-stage model of category learning consisting of formation of a task-independent shape-selective representation that then provides input to circuits that learn the categorization task. To test this model, we trained human subjects to categorize car images generated with a morphing system that allowed us to precisely control stimulus shape and define a category boundary. We scanned the subjects before and after training using fMRI rapid adaptation techniques to explore shape- and category-selective neural plasticity induced by categorization training. By scanning trained subjects while they performed different tasks we further explored how the learned representation was modulated dynamically by task demands. Last year we reported preliminary results supporting the two-stage model – with increased shape selectivity in the LOC region (the human homolog of monkey area IT) and a weak learned category selectivity in prefrontal regions. We have now trained and tested more subjects, ran additional experiments, and completed the data analysis. The new results confirm and strengthen the earlier findings. Furthermore, there are several additional observations: (1) category learning sharpened neuronal shape selectivity in LO but did not induce any category selectivity, suggesting that learning in these areas is driven predominantly by bottom-up signals; (2) strong category selectivity was found in PFC (with a right hemisphere dominance), and this category selectivity could be totally dissociated from shape change; (3) the category-selective activation in PFC was not only task-dependent (with the region not showing any category selectivity when subjects performed a task for which the learned categories were irrelevant) but also showed a strong correlation with subjects’ performance; (4) category training also improved subjects’ ability to discriminate the stimuli, yet the activations of “core” FFA neurons did show neither shape nor category selectivity for cars, arguing against theories that have postulated a central role for the FFA in “expertise processing”. The present study thus suggests that category learning in humans and monkeys (Freedman et al., 2001, 2002, 2003, 2005) follow similar principles. Finally, the results show that fMRI-RA techniques can be used to investigate learning effects at a more direct level than conventional analyses based on comparing average BOLD-contrast response amplitude in response to individual conditions, providing a powerful tool to study cortical plasticity and task-related effects in humans.

Talks for NIH-Conte Meeting – September 12, 2006

The Role of the Frontal Eye Field in the Selection of Representations in the Visual Cortex

Tirin Moore
Massachusetts Institute of Technology

Navigating Cluttered Acoustic Environments

Shihab Shamma
University of Maryland

Investigating the Role of Top Down Contextual Priors on Object Search in Real World Scenes

Aude Oliva
Massachusetts Institute of Technology

Behavioral studies have shown that human observers make extensive use of contextual scene information during object search in natural images. In a series of experiments, we investigated the influence of top-down contextual priors in object search by monitoring eye movements as participants searched real-world scenes for objects (e.g. a pedestrian, a man or a woman, a painting, a cup). In our first study, we show that a computational model that relies on top-down categorical priors (the identification of the scene as a street, a park, etc.) can predict the location of human eye movements when the target object is small and camouflaged in the scene. In a second study, we investigate the relative contributions of bottom-up saliency and top-down contextual priors in predicting eye fixations during object search. Finally, in a third study, we evaluate how the strength of contextual priors (e.g. the probability of the association between a scene context and the object presence or location) influences the different stages of object search: the initial glance at the scene, the search process itself and the processing of the target object. Altogether, the results suggest a key role for experience-dependent top-down influences during visual search and impose constraints on computational models of vision. Work in collaboration with Barbara Hidalgo-Sotelo, Naomi Kenner, Antonio Torralba.