MIT Progress Report Summary - 2004

1. Specific Aims

The specific MIT project, led by Dr. Poggio, involves aims both in computation as well as in physiology in monkeys. Computational modeling of visual cortex interacts daily with experiments in the physiology labs of E. Miller and J. DiCarlo. Our aim was – and is – to be guided by the model of recognition, probing the relations between identification and categorization and the properties of selectivity and invariance of recognition, especially with image clutter, in IT and PFC cortex of behaving macaque monkeys.

In particular, DiCarlo, Riesenhuber and Poggio are beginning to test the hypothesis that selectivity and invariance of IT neurons depends on clutter during training and testing. In the second part of this project Miller, Poggio and Riesenhuber, are investigating the neural bases of different recognition tasks, *identification and categorization,* and their interaction in perception, specifically in the phenomenon known as Categorical Perception. In addition, Poggio 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 is developing into a close collaboration with CalTech on the computational side and with Northwestern on the experimental side, possibly including new experiments (not in our original proposal) about tuning in V1. Our specific aims are listed below:

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

Aim A.2: *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 two new collaborative aims: 1) in addition to exploring as planned (with CalTech and Northwestern) the biophysical mechanisms underlying the max operation, we now plan to also explore the mechanisms underlying the Gaussian-like tuning of cortical cells; 2) we started collaborating with the Harvard lab of Dr. Livingstone on properties of V4 cells: their experiments with 2 and 4 spots allow to test predictions of the model and refine its architecture and parameters.

2. Studies and Results

- **"New" Aim N.1:** We had planned to study computationally (with CalTech) and experimentally (with Northwestern) the circuitry underlying the max operation. In the meantime a paper is ready for publication. We also started to work on a new related and equally interesting question (see above): what is the basis for the other key operation in the model, eg Gaussian-like, multidimensional tuning of cortical cells? The preliminary answer from initial modeling work is that it may be a very similar circuit to the one postulated for the max operation. We are working on the hypothesis that Gaussian-like, multidimensional tuning – as found in many neurons in cortex – can be generated by *normalization* of the input vector, followed by a simple threshold-like sigmoidal nonlinearity. In fact this may be the main reason for the widespread presence of gain control circuits in cortex, where tuning to optimal stimuli is a common property. We will explore a specific circuit for normalization, based on lateral shunting inhibition. The same basic circuit of lateral inhibition could underlie Gaussian-like tuning (via normalization) *and* the Softmax operation – which are the two key operations required at various stages in the model of object recognition (*Knoblich, Poggio*).
- **"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 (such as spare spot stimuli and contour features). This effort will help to make the model more biophysically plausible and has already produced several interesting new ideas about the underlying neural circuits in V4 (*Kouh, Livingstone, Freiwald, Poggio*).

- **Aim A1:** We have made significant progress on this aim. The first step is to train monkey subjects to be experts at detecting specific objects. One monkey subject has been trained in a *sequential object recognition task* that requires the detection of a specific shape (the *target shape*) embedded in a temporal sequence of shapes drawn from the same, parameterized shape space (the *distractors*). To insure the generality of our results, the monkey has been trained to detect a target object in each of three different parameterized shape spaces (cars, faces, and abstract silhouettes). During training, an adaptive staircase has been used to control the similarity between the distractors and the target shape according to the behavioral performance of the animal. This forces the animal to detect more and more subtle differences between the target shape and the distractors (become more of an expert in each shape class), and has allowed us to quantitatively measure the animal's level of expertise across training days (threshold). Results of the training in the first two shape spaces each showed a consistent performance improvement (more than doubling) during the first 7-10 days of training that reached an asymptotic value that remained constant for the remaining 8-10 training sessions. Based on previous studies, it is expected that this training has resulted in IT neurons with shape selectivity within each of the shape spaces. An important question that we will be able to answer is: does such training result in more IT neurons tuned to each of the target objects or is the IT population tuned evenly across the shape spaces? In either case, IT neurons tuned to objects within one or more of these shape spaces will be the focus of our neurophysiological work on clutter (*Zoccolan, DiCarlo*).
- **Aim A2:** The tuning properties of the shape-tuned neurons in clutter will be measured by the simultaneous presentation of the optimal and sub-optimal "cluttering" stimuli of varying similarities, which will be from the same and the different shape spaces as the optimal stimulus (*Zoccolan, DiCarlo*). We have investigated the learning of intermediate features (in particular in the model layers corresponding to visual area V4) to improve object detection in cluttered scenes. We find that unsupervised learning of features greatly improves the system's ability to detect object from the target class in natural scenes. Moreover, the biological architecture appears to have advantageous computational properties in terms of transfer of learning across tasks, relevant for subproject B (*Serre, Riesenhuber, Poggio*).
- **Aim B1:** We examined a population of 144 PFC neurons and 151 ITC neurons in the lateral prefrontal cortex (PFC), while monkeys alternated between categorizing our morph stimuli into "cats" versus "dogs" and matching specific, individual, category members. We found that the activity of some PFC neurons encoded category, others the specific individuals, and further, that the majority of neurons showed similar activity across the two tasks. This suggests common substrates for each task, or perhaps adoption of a "hybrid" strategy in which the monkeys simultaneously encode the category membership, as well as the identity of individuals (*Freedman, Serre, Riesenhuber, Poggio, Miller*).
- **Aim B2:** Monkeys are currently being prepared for this aim (*Roy, Miller*).

3. Plans

- **Aim A1:** We will investigate the clutter tolerance properties of IT neurons in the first monkey beginning in June, by performing electrophysiological recordings from the IT cortex. The neuronal recording protocol has been designed to specifically assess: 1) the distribution of IT tuning to target and non-target objects in these spaces; 2) the sharpness of that shape selectivity; 3) the position tolerance of that shape selectivity; 4) the amount of interference (neuronal response reduction) caused by shapes flanking the neuron's preferred shape (inside the region of position tolerance) as a function of the similarity of the flanking shape to the preferred shape. Most of these recordings will be performed during passive viewing to avoid training the monkey in clutter conditions, and to allow us to rapidly test many stimulus conditions. We expect this recording to be largely complete in the first animal by the end of Year 1 of the project.
- **Aim A2:** We still plan to train the animal to detect the same target shapes in the presence of flanking distractor shapes (clutter). This training will be followed by a re-assessment of the same IT neuronal properties described above. We expect this to be complete in the first animal by the end of Year 2 of the project. Depending on the results of the first animal, a second monkey will begin training in Year 2 to replicate all findings. We plan to have completed analyses and begin submitting publications in Year 3. Simulations of plausible circuits for tuning and for Softmax.

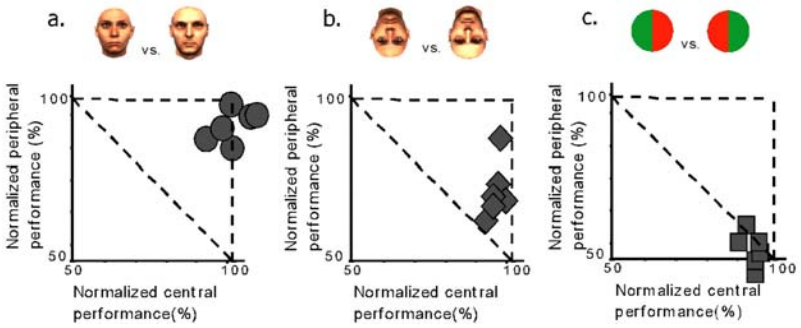
Caltech Progress Report Summary

1. Specific Aims

The Caltech project is organized around the central theme of attentional aspects of object recognition, using human psychophysics and computational and biophysical modeling. The research is organized into three aims: (i) Psychophysics of attention and recognition in natural scenes will parallel electro-physiological work in monkeys (DiCarlo) and determine the effects of "physical" distance between stimuli (*clutter*) and "similarity" distance between targets and distractors *(task complexity*) on a task's attentional requirements, using both familiar and unfamiliar stimulus categories. This will help determine the limits of the current MIT feed-forward recognition and saliency models. (ii) In light of these constraints, the saliency model previously developed in Koch's group will be integrated with the feed-forward recognition system to implement attentional modulation of object recognition. (iii) Finally, the Koch group will evaluate plausible network and single neuron models of how the MAX operation could be carried out in cortex, aiming to guide the electrophysiology and account for the results obtained by the Ferster group.

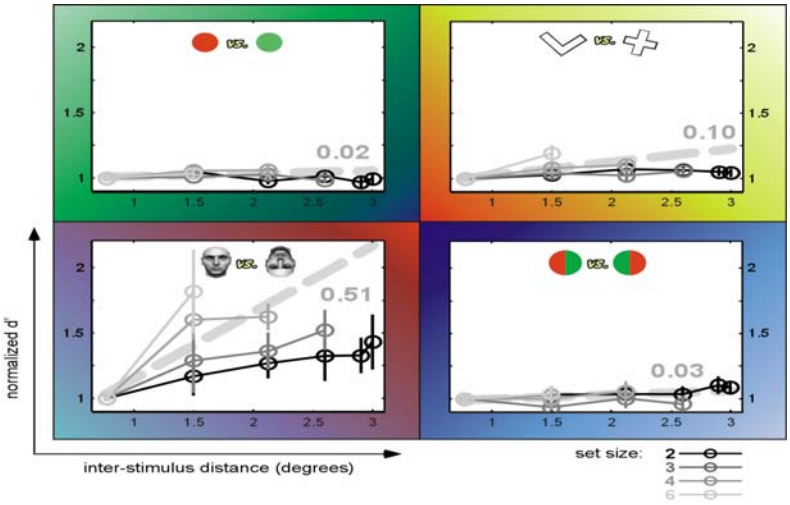
2. Studies and Results

Face-Gender Discrimination is Possible in the Near-Absence of Attention. Simple visual tasks such as orientation or color discrimination can be performed in the near-absence of spatial attention. In contrast, participants are unable to perform slightly more complex tasks, such as discriminating between the arbitrarily rotated letters "T" and "L. However, it has been recently shown that natural scenes (e.g., animal vs. non-animal) can be categorized in the near-absence of spatial attention (Li *et al.,* 2002, Rousselet *et al*., 2002). Where does this ability to process natural stimuli in the near-absence of spatial attention break down? To answer this question, we chose a task that involved a fine discrimination of the spatial arrangement of features that are present in both targets and distractors. We tested 6 subjects on a face-gender discrimination task using the dual-task paradigm. Our results show that performance on this task suffers only minimally when attention is unavailable. Additionally, this high-level of performance is not due to low-level differences between male and female stimuli used – when subjects do this task on a set of inverted faces, dual-task performance drops significantly.



Further, performance on another task known to require attention falls to chance levels indicating that the attentionally-demanding task in our dual-task paradigm efficiently engaged spatial attention (L. Reddy, P. Wilken, C. Koch).

Clutter effects in visual search: human psychophysics. Recent experiments from our lab revealed that categorization of an isolated natural scene or object could be performed in the near absence of attention (Li *et al.*, 2002; Reddy *et al.*, 2004). Surprisingly, the same tasks were performed poorly in visual search conditions where a target must be detected among multiple distractors. We have proposed that for these natural scenes and objects, parallel visual search performance – the hallmark of pre-attentive processing – might be impaired by receptive fields clutter (VanRullen *et al.,* 2004). To demonstrate this, we used a modification of the visual search paradigm in which we varied parametrically the set size (number of elements presented) and the inter-stimulus distance (an inverse measure of clutter). We found that, only in the case of a task involving natural target objects (face photographs), decreasing clutter could yield significant performance improvements in visual search (in fact, sensitivity was found to increase by more than 50% for each 1º increase in inter-stimulus distance).



This result is a significant first step in bridging the gap between attentional requirements recorded on isolated stimuli (dual-task) and those estimated from cluttered displays (visual search). Further, it provides a human psychophysics counterpart to the electrophysiological experiments of DiCarlo, investigating clutter effects in visual cortex (L. Reddy, R. VanRullen, C. Koch).

Biologically plausible face detection in cluttered scenes

We previously demonstrated that endowing the biologically plausible HMAX object detection system with feature learning at the intermediate S2 level enables the system to successfully detect faces in images (Serre *et al*. 2002, Louie 2003). While this system is robust to small shifts or changes in scale, it is by itself not able to pick out and, say, count the faces in a cluttered scene. We have also demonstrated that our saliency-based attention model, extended for the selection of salient regions, can facilitate learning and recognition of multiple objects in cluttered natural scenes (Walther *et al.* 2004, Rutishauser *et al.* 2004). In a natural extension of these two lines of work, we are combining the two efforts to yield a biologically plausible vision system capable of detecting faces in highly cluttered scenes and with potentially many target objects present. First explorations of the combined system provide encouraging results. In 500 test scenes created by pasting between 1 and 9 synthesized face images into cluttered background images, the combined system was able to detect 95.8 % of the faces with false alarms in about 10 % of the images. For the final stage of the detection in HMAX, we are using a support vector machine (SVM) classifier as well as a biologically more plausible nearest neighbor (NN) classifier. The performance of the two classifiers is comparable, with the NN classifier showing on average 5% lower detection rates than the SVM. Our next step is to train the detection system on real natural scenes with faces. We are currently in the process of collecting a training and test data base of images for this purpose (D. Walther, T. Serre from MIT, C. Koch).

3. Plans

In agreement with our earlier proposal, we will continue to investigate both psychophysically as well as computationally the interactions of selective visual attention with recognition. Given the availability of a 3.0T human fMRI scanner at Caltech, we are planning on repeating some of the attentional dual-task experiments reported above in the scanner to assess the extent to which the behavioral response correlates with attentional modulation in relevant cortical areas (as measured using fMRI BOLD). In particular, we will focus on the Fusiform Face Area (FFA) in human subjects to measure its BOLD response under three conditions (with the same physical input): the subject pays attention to the central, letter task, the subject pays attention to the peripheral, face-gender task or the subject performs both task. We know that the central, letter task is attentional-demanding. Will this show up in a reduced BOLD respond in the FFA, even though the subject can still carry out the face-gender discrimination task at a high level of performance?

Georgetown Progress Report Summary

1. Specific Aims

The Georgetown subcontract was originally part of the MIT project, the new subcontract being necessitated by Dr. Riesenhuber's move to Georgetown University to start a lab there. Thus, the specific aims of the Georgetown subcontract directly follow from Riesenhuber's role in the MIT project. In particular, Riesenhuber, together with Dr. Poggio, directs the modeling effort of the project, and the integration of model simulations and experiments. The projects therefore are tightly integrated with the MIT effort. Particular aims of the original MIT project we have been focusing on are:

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

Aim A.2: 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.

In addition, we have also been closely involved with the Northwestern project (see below), in particular their aim 1 ("Test whether a subset of the complex cells shows a MAX-like pooling of inputs of the simple cell type"). Finally, we have started to investigate possible mechanisms of attentional modulation through model simulations, relevant to the Caltech part of the proposal.

2. Studies and Results

Aim A.1: We have analyzed in the model the interaction of two stimuli in the receptive field of a model IT unit of a monkey trained to recognize individual novel stimuli, to provide predictions for DiCarlo's experiment. In particular, we have been focusing on the model prediction of a U-shaped dependency of the neural response as a function of the similarity of two simultaneously presented stimuli, one being the cell's preferred (and trained) stimulus, and the other a stimulus of varying physical similarity to the preferred stimulus (see Figure 13 in the proposal). We found that the strength of the "U-effect" depends on the specificity of the features of units (e.g., in V4) providing input to the IT units. To develop the paradigm and stimuli for DiCarlo's monkey studies, we have further conducted psychophysics with human subjects (funded and conducted separately – we have included funding for future experiments as part of the Georgetown subcontract). This has been helpful in determining which target stimuli and distracters to use for the monkey experiment. Very interestingly, we found support for a U-shaped interaction also in the human behavior, as predicted by the model. This effect was much weaker for the familiar object class of faces. This result can be understood with the assumption that faces are represented by a population code over "face units" rather than by individual "grandmother units" (which are hypothesized to be learned for the novel stimuli). We are currently investigating this hypothesis further.

Aim B.1: Together with Thomas Serre, we have been analyzing the physiological data obtained by David Freedman in Earl Miller's lab from two monkeys trained to switch between a categorization and a discrimination task on the cat/dog stimulus space. One focus of the analysis is to test the model hypothesis regarding the location and timing of task-specific modulations of neural responses. One problem with the discrimination/categorization design is that there are no clear tests to determine whether a neuron is involved in the discrimination task, unlike in the categorization task, where one can test for a neural correlate of the category boundary. We have thus decided (together with Earl Miller and postdoc Jefferson Roy in Miller's lab) to develop a training paradigm and stimuli to train a monkey to switch between two different categorization schemes on a subset of the cat/dog morph space. This design has

the advantage that, based on our previous work, we can expect clear neuronal correlates for the individual tasks, which will enable us to better understand the neural bases of different recognition tasks. The monkey has now been trained and is being prepared for neurophysiological recordings.

Northwestern Project: We have finished the analysis of the complex cell data obtained in Ferster's lab, finding that the response of some complex cells can in cat striate cortex be well described by a MAX-like function. A paper on these results is under review (Lampl, Ferster, Poggio, and Riesenhuber, "Spatial Integration and the MAX Operation in Complex Cells of the Cat Primary Visual Cortex Revealed by Intracellular Recordings").

Caltech Project: We have investigated through model simulations under what conditions featural attention can improve object recognition performance (*Schneider and Riesenhuber, 2004*). The simulation results suggest that, unlike spatial attention, featural attention can improve object recognition performance only in very limited circumstances.

3. Significance

We refer to the section in the Center part of the report.

4. Plans

Aim A.1: After using model simulations and human psychophysics to inform the monkey training paradigm, we will next perform simulations to provide quantitative predictions for the tuning properties of IT neurons involved in the task (see the MIT report). These predictions will be compared to actual IT neuron tuning as data from DiCarlo's lab becomes available.

Aim A.2: We next plan to study how IT neuron tuning properties would be expected to change in the model as a result of training in clutter. These simulations will be developed in conjunction with human psychophysical tests to optimize the training regime in clutter for the monkey training.

Aims B.1 & B.2: We will continue the analysis of the existing data to determine whether we can find evidence for task-dependent modulations in subpopulations of neurons in IT and PFC. Once physiological data from the monkey trained on the category-switch task becomes available, we will compare these data to the model predictions. In parallel, we will also start developing the training paradigm for the Categorical Perception task.

Northwestern Progress Report Summary

In our previous work, we have explored the MAX operation in complex cells of primary visual cortex in the cat, asking whether or not complex cells actually perform a MAX-like computation on their inputs. We did so by presenting bar stimuli in pairs within the receptive fields of the complex cells and comparing the response to the paired stimuli, R_{a+b}, to the responses to each stimulus presented alone, R_a and R_b. In some cells R_{a+b} was equal to the sum of R_a and R_b. That is, the complex cells summed its visual inputs linearly. In other cells, R_{a+b} was greater that the sum of R_a and R_b (facilitation, or supra-linear interactions). But in the majority of cells R_{a+b}, was less than the sum of R_a and R_b. And in particular, R_{a+b} was similar to the larger of the two individual responses, R_a and R_b. That is, these cells behaved in a MAX-like way. To quantify this behavior, we calculated a MAX index

$$I = (R_{a+b} - MAX(R_a,R_b))/MIN(R_a,R_b),$$

which is near 0 for cells with MAX-like behavior.

Movshon *et al.* (1978) observed two types of interactions between bars flashed simultaneously in complex receptive fields. When two bars of the same polarity were presented nearly adjacent to one another, the response was greater than either of the individual responses, but less than their sum (MAX index greater than 0, but significantly less than 1). When bars of opposite polarity were presented nearly adjacent to one another, the response was less than either of the individual responses (MAX index less than 0). When the bars were presented farther apart from one another, the opposite behavior was observed, with MAX indices less than 0 for the same polarity and greater than 0 for the opposite polarity. Effects like these are expected from the feedforward model of Hubel and Wiesel in which complex cells receive input from multiple simple cells. Closely-spaced pairs of bars will fall into the same subfield of a presynaptic simple cell. When of the same polarity, they will facilitate one another; when of the opposite polarity, they will antagonize one another, independently of the pooling mechanism used by the complex cell. More widely space bars will fall into different subfields of a simple cell afferent and so will facilitate one another when of opposite polarity and antagonize one another when of the same polarity.

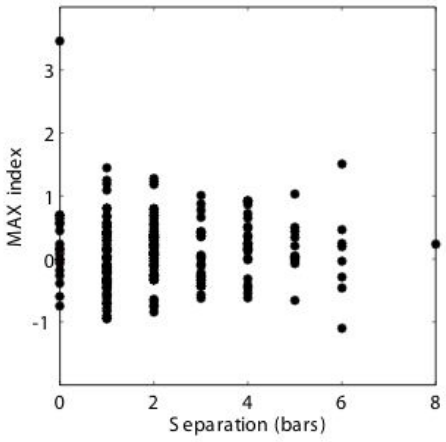


Figure 1. *MAX index plotted against separation between the two stimuli in a pair (measured in bar widths orthogonal to the preferred orientation of the cell) for bar pairs with the same polarity (A, n = 214), and for pairs in which one bar was dark and one bright (B, n=50).*

Strikingly, we did not observe the pattern of behavior described by Movshon *et al.* (Figure 1). Instead, we found that the interaction between pairs of stimuli was almost completely independent of bar polarity and the distance of separation between bars. Our investigation of these two variables, however, was not systematic. The location and polarity of bar pairs were selected to maximize the amplitude of the responses, and only a few bar locations were explored in each cell. The trend we observed (or lack of trends) comes from data pooled across many cells, which could obscure any real dependence.

This discrepancy between Movshon *et al.*'s data and ours is an important issue for the MAX operation. The MAX operation is proposed to be independent across the receptive field, and yet the Movshon *et al.* paper, which shows a strong dependence on pair position, is a classic in the field and widely cited as evidence regarding the mechanisms of complex cell receptive field construction. We are therefore pursuing this issue experimentally in two ways.

First, we are presenting pairs of bars in complex and simple receptive fields using a complete stimulus set. After mapping the receptive field, it is divided into 8 or more bar-shaped regions, and pairs of stimuli for all possible pair-wise combinations of the bars are presented in rapid succession in random order. The complete set of pairs is presented repeatedly, in a different random order each time. In this way, within a single cell we can explore the systematic relationship between inter-bar distance and pair-wise interactions. So far, in the very few cells we have examined, our data are coming out similar to what we observed previously.

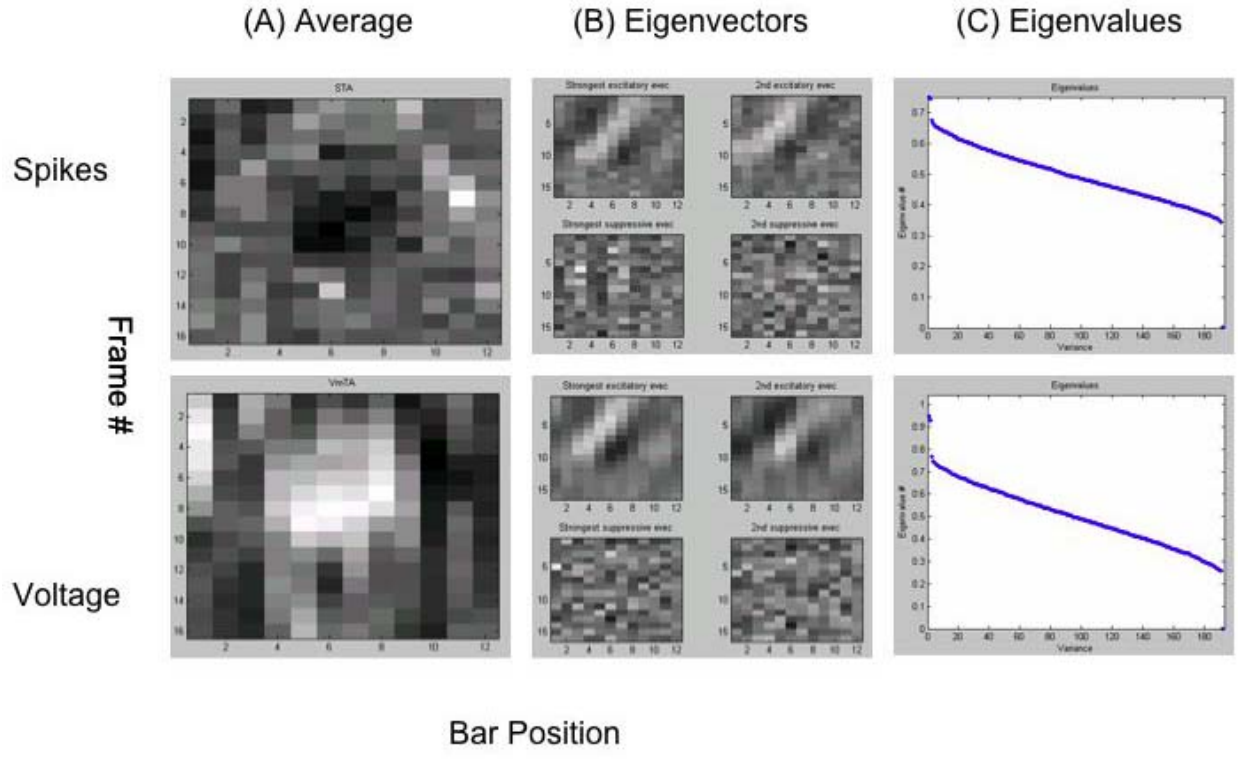


Figure 2.

Second, we are applying the bar response data to a novel analysis of receptive fields developed by Simoncelli *et al*. In this calculation, the data are cross correlated with the randomly presented stimuli and then applied to a singular value decomposition. Significant 2-D eigenvectors are taken to be components of the receptive fields. In the case of complex cells, these components are likely related to the underlying inputs to the cells. An example is shown in Figure 2. In A, the standard, stimulus-triggered average of the responses shows the receptive field as it is commonly described (spike-defined responses above and sub-threshold membrane potential responses below). In B, the significant eigenvectors are shown. These resemble classical simple cell receptive fields in a spatial quadrature relationship. We will be exploring the significance of these component responses and how they relate to the pair-wise bar interactions from the previous experiment and how they relate to the MAX calculation in simple cells.