Chapter 9: Conclusions and Future Work

The planning and execution of grasps requires the integration of multiple sources of information. The visual system provides the object's identity, as well as an estimate of its shape, size, and localization in space. Although this information can be used to specify the fine details of a grasp plan, a single object actually affords many different grasps. This choice of grasp is dependent upon the behavioral context, which can include instructions from an experimenter, working memory inputs, or a specification of the task that is to be performed with the object once it is grasped.

This thesis has presented a neural-level model of the cortical processes that are involved in the generation and execution of grasp plans. The model addresses the issues of neural-level encoding of object, grasp, and phasic information, and shows how populations of cells can capture the critical information that is required to perform the necessary computations. In addition to reproducing single-cell behavior that has been observed experimentally in F5 and AIP in monkey, the model makes several more general hypotheses about the computations that are being performed by these neural regions:

1. The neural activity patterns observed in AIP reflect the extraction of action-relevant features (or affordances) from the visual representation of the attended object.
2. During execution of the grasp, AIP maintains an *active memory* of the affordance that corresponds to the grasp that is being executed. This active memory is maintained in part by recurrent projections from F5.

3. During grasp plan formation, F5 selects from the possible grasps, as represented in working memory or in AIP (corresponding to the affordances). This selection process takes into account task-related constraints (received from either area 46 or F6).
4. F5 is responsible for the high-level execution and monitoring of the grasp.
5. Parameters of an object or grasp are reflected in sub-populations of F5 and AIP cells (this hypothesis is explored in Prediction #3 of Chapter 6).

These hypotheses lead to a number of predictions about the single cell- and population-level behaviors under several novel tasks.

1. In the case where two distinct objects are presented and subsequently grasped in an identical manner, hypotheses 1 and 2 imply that motor-related cells in AIP will tend not to reflect differences in the two situations. However, visual-related cells (which derive many of their inputs from PIP and IT) can reflect these differences, and hence appear to demonstrate an object-specific code (Prediction #1 of Chapter 6).
2. Hypothesis 1 implies that different parameters of an object will be inequitably reflected in AIP. Specifically, we expect to see only those parameters that are relevant for programming of the grasp (Prediction #5).
3. In the case where an object is to be grasped in one of two ways, but information about which of the two grasps that must be selected (e.g. via an instruction stimulus) is withheld for some delay period, hypothesis 3 implies that we would expect both possible grasps to initially be prepared, and hence we predict activity patterns in F5 that correspond to both possibilities. When the IS is given, the population of cells not corresponding to the selected grasp are inhibited (Prediction #2). Hypothesis 2 implies that AIP will behave in a similar manner.
4. Because F5 monitors the ongoing grasp (hypothesis 4), when there is a miscalibration between the visually-estimated and actual size of an object, we predict a shift in the executing program at the time that the more precise tactile information becomes

available (Prediction #4). Hypothesis 2 implies that this shift in program will also be reflected in AIP.

As discussed at the end of Chapter 6, some observations made of AIP cell responses are consistent with Predictions #1 and #2.

PET Imaging

In addition to the predictions made by the model at the neuronal level, we have shown how it is possible to extract other measures of model behavior. Specifically, through the application of the synthetic PET imaging technique (Arbib, et al., 1995), in which the regional synaptic activity is measured over the course of performing a task, we are able to make predictions about the expected results from human PET experiments.

Two sets of predictions made by the model were tested in a human PET experiment. In the first experiment, we looked at the relative synaptic activity during execution of a conditional task (in which an IS informed the subject as to which grasp to use) as compared to execution of the two grasps without using an IS (see Chapter 7 for more details). The most significant result from this comparison was that an area of the contralateral postcentral gyrus (corresponding to dorsal premotor cortex, or F2) was very active. This result was predicted by the model. In addition, the model predicted a reasonable activity increase in F5, and a slight increase in AIP (increase from non-conditional to the conditional task). The human PET result did not confirm these predictions: no change in activity was seen in F5, and only a small amount of activity was observed in AIP.

The second of the comparisons looked at synaptic activity differences between the precision and power grasps (in the model, we compared precision pinch and side opposition). The model predicted an increase in both F5 and AIP (from side opposition to precision pinch). The human PET result showed an increase in AIP activity, but no change in F5.

We concluded the discussion on PET and synthetic PET imaging by showing how the PET results can be used to further constrain the model. Because PET imaging inherently provides a global measure of system behavior, the constraints that it provides are also at a global level. These global constraints allow us to make conclusions about the existence or non-existence of projections from one layer to another, or even about the relative strengths of inhibitory and excitatory projections between regions.

Grasp Configuration Learning

The primary model in this thesis relied on a network that was, for the most part, wired by hand. However, in Chapter 8, we demonstrate how the problem of transforming visual and task parameters into an appropriate grasp description can be learned without prewiring a network. The transformation was learned in a trial-and-error fashion: for a given situation, the network produced a grasp plan and then observed the outcome of the plan. Using a reinforcement-learning algorithm, the success and efficiency of the executed grasp was used to update the transformation. Most interestingly, after learning was complete, some neurons within the network became sensitive to the type of grasp that was performed (precision pinch or power grasp), as we have seen in the monkey F5 (Rizzolatti, et al., 1988). In addition, other cells reflected the aperture width of the grasp. In addition, the behavior of some cells could not be described using a simple symbolic rule, but demonstrated rather complex behaviors (as we see in the biology).

Related and Future Work

Control of Wrist Orientation

Some cells observed in AIP are responsive to the orientation of the target object (Taira, et al., 1990). Although this information might be relevant to the set of grasps that are afforded by the object, it also likely that this information is used in programming of the wrist position.

Note that the necessary wrist position is not simply determined by the orientation of the target object in space, but varies as a function of the affordance that is ultimately selected. Including object orientation in AIP is consistent with the notion that F5 is involved in controlling the position of the wrist (Rizzolatti, 1995).

Planning of Arm Movements

A related problem is that of positioning of the arm. Although the location of the arm is tied to the position of the object in space (as represented in VIP), depending upon the grasp that is selected, the wrist might be positioned differently (consider grasping a cylinder with either a precision pinch or a power grasp). Thus, there must be some mechanism by which the selection of an affordance/grasp pair is integrated into the goal position of the arm. Likewise, this mechanism must also permit arm position constraints to be reflected in the decision of the grasp. Possible pathways for these interactions include AIP's connections with VIP and F4.

The transformation of object position from a body-centered coordinate system into a program for arm movement has been studied extensively (Kalaska & Crammond, 1992; Stein, 1992; Wolpert, Ghahramani, & Jordan, 1995). In particular, (Burnod, Grandguillaume, Otto, Ferraina, Johnson, & Caminiti, 1992) have proposed a model based on monkey neurophysiological studies of visual and motor encoding of movements. The model learns to combine a visual representation of target position and a kinesthetic representation of the current arm state in order to compute arm movement commands. The movements are represented using a population coding scheme.

Coordination of Reach and Grasp

During execution of the reach and grasp, it has been observed in human that there is a close temporal coordination between the reach and grasp processes (Paulignan, et al., 1991a; Paulignan, MacKenzie, Marteniuk, & Jeannerod, 1991b). (Hoff & Arbib, 1993; Hoff, 1992) have proposed a schema-level model that describes how the reach, preshape, and enclose modules exchange information in order to achieve this coordination. During execution, each module estimates the amount of time that is required in order to complete its goal. This timing information is compared by a coordinating module, which in turn modulates the rate of execution for each module. In cases where the reach process requires much more time than the preshape/enclose processes (as in the case where the target position is perturbed), their state is reset to the beginning of the preshape phase. Although this coordination mechanism most likely involves cerebellar processes, the cerebellum has access to the hand primarily through F5 and MI. Within the model presented in this thesis, modulation of the 'rate' of execution can be achieved through a global facilitation or depression of F5 or MI activity. The resetting of hand control state could be achieved through the shunting of flexion phase cells (in F5) and activation of the extension phase cells.

Learning of Motor Programs

In Chapter 8, we presented a second model in which the mapping of visual and task parameters to a selected grasp was learned through a trial-and-error procedure. One of the significant simplifications of the model was that the mapping was static -- i.e., the grasp plan that was output was independent of the module that executed the plan. Although it is not immediately clear how the set of grasp programs in F5 can be learned, once the set of programs exist, it is possible for the mapping between visual parameters (as represented in PIP and IT) to be associated with the affordance representation within AIP. This learning scheme relies on two features in the model. First, that F5 will automatically adjust its program to the actual situation based on tactile and kinesthetic inputs received via SII. Second, changes in the F5 program are reported back to AIP through the recurrent cortico-cortical projections. Thus, once the grasp is established, AIP contains a representation of the grasp that was *actually executed*, which can then be associated with the visual representation of the object.

Encoding of Grasp Programs in F5

As we observed in Chapter 6, individual F5 units in the model were often active in more than one phase of the grasping program to which they belong. In addition, a number of cells were observed as participating in more than one grasp. Despite the temporal and programmatic overlap, as implemented, the model does not make functional use of this feature. By *functional use* we mean that the output motor command generated by a single F5 unit (to MI) is specific to one grasp type/phase pair [1]. A more satisfying grasp coding scheme, and one that is better related to Rizzolatti's notion of a *language of motor acts* (Rizzolatti, et al., 1988), is one which allows individual F5 units to be responsible for small pieces of a grasp, which can then be assembled with other pieces to form a complete grasp. For example, an F5 unit that participates in both a lateral and precision pinch might generate motor outputs that are common to both grasps, such as the control of thumb flexion. Although it is possible to approach this issue through modification of the network wiring heuristics (see Appendix A), it is probably best addressed in the context of learning the motor programs.

Perception of Grasp

Rizzolatti, Fadiga, Gallese, & Fogassi (1996) have recently reported that many neurons in the convexity of area F5 are not only active as the monkey executes a specific grasp, but also when the monkey observes the same motor act being executed by the experimenter or another monkey. It has thus been hypothesized that these neurons are involved in the perception of motor acts. These observations present a number of interesting challenges for our modeling work. First of all, how do these neural properties arise? One possibility is that the monkey first learns to associate the visual patterns created by movements of his own hand (possibly supplied by temporal cortex) with the descending motor command produced by the F5 neurons

[1] However this overlap does play a role in the cooperative interaction of groups of F5 units.

that are responsible for the generation of the movement (which are located in the bank of F5). These associations are then generalized to hand movements external to the monkey. Second, how is this information useful to the monkey? One important reason is that through understanding the motor acts that are being performed around him, it is possible for the monkey to learn when and how to perform the same motor acts. This ability to learn by example of others is also important for our work in the design of more intelligent robot controllers (Fagg, 1993; Kang & Ikeuchi, 1995).