Chapter 4: A Region-Level View of the Reaching and Grasping Process

This chapter outlines the computational and neurobiological constraints for our model of primate reaching and grasping. From these constraints, we outline a region-level view of the model, and show how the different brain regions interact to perform a grasping task.

Within the model, the dorsal and ventral visual streams first compute a representation of the object of interest. We hypothesize here that information from both streams converge on AIP, where the set of appropriate grasps (or affordances) is computed. These grasp-centered codes are then passed to area F5. We propose that it is this region that is responsible for first selecting from one of the possible grasps, and then unfolds this representation in time to execute the preshape/grasp motor program. During the execution process, AIP performs as an active memory by maintaining a representation of the grasp that is actually executed.

In the subsequent chapters, we will discuss the neural network implementation of the model, and then demonstrate its behavior in simulations of several experimental contexts.

The Reaching/Grasping Task

The reaching and grasping tasks studied in this work are derived from the protocols used in the laboratories of Sakata and Rizzolatti (Rizzolatti, Camarda, Fogassi, Gentilucci, Luppino, & Matelli, 1988; Taira, Mine, Georgopoulos, Murata, & Sakata, 1990). In these tasks, the monkey sits directly in front of a touch pad or a switch that he must engage with his hand. An object of some shape and dimension, selected by the experimenter, is positioned just within reach - directly in front of the monkey. The following outlines the sequence of events that take place during one such protocol (of Sakata):

Step 1: The monkey, sitting in the monkey chair, focuses his attention on an LED.

Step 2: A ready signal (turning on of the LED) indicates the start of a trial. The monkey responds by placing his hand on a touch pad and fixating on the LED, which will indicate when movement may begin. The LED is positioned (using a horizontal half-mirror configuration) such that it appears superimposed on top of the object to be grasped. The object is lighted in such a way that it is also visible.

Step 3: The LED changes color (the *GO signal*). The monkey responds by removing his hand from the touch pad and reaching towards the target object. During the reaching phase of movement, the monkey preshapes his hand in anticipation of making contact with the object.

Step 4: Contact with the object is made, and the monkey secures a grasp. A manipulatory movement is then made, such as pulling on the object or pushing it in some specific direction. The resulting position is held for a random delay period.

Step 5: The LED changes color again (the *secondary GO signal*). The monkey responds by releasing the object and moving his hand away.

If all steps of the task are performed properly, the monkey is rewarded (in this case with a drop of juice), after which the next trial is begun.

For the most recent set of experiments performed by Sakata, objects have been one of: cylinder, plate, torus, cone, sphere, and cube. The dimensions of these objects are also varied. Earlier experiments have made use of objects such as joysticks, buttons, and pull-knobs.

Model Goals

Even the simple task presented here provides a number of interesting challenges that we approach in the model below. These include:

1. What mechanisms are involved in the transformation of a visual description of an object into a grasp that is appropriate for interacting with that object?

2. How may a variety of sources of information be used to select a single appropriate grasp? This information includes visual (from both the how and what pathways), context-related (what is the current task?), and abstract stimuli (e.g., an instruction stimulus).

3. How can the decision to execute a particular sub-program modulate decisions made for other components of the program? For example, how does the selection of the grasp affect the positioning of the wrist, and how can we implement this computation neurally?

4. How is the program coded at different levels? In AIP, what are the primary parameters being represented; how are they coded? In F5, how are the grasps represented and then played out further down stream?

5. What mechanisms provide the neural activity patterns seen experimentally, especially within AIP and F5?

Development of the Reaching/Grasping Model

Visual Processing (The How, What and Where of Objects in Space)

When a human or monkey directs attention towards an object, what types of information are extracted from the resulting retinal image that ultimately will be useful in somehow interacting with that object? Goodale and Milner have argued that two very different classes of information are extracted by distinct visual pathways (Goodale, Meenan, Bulthoff, Nicolle, Murphy, & Racicot, 1994b; Grady, Haxby, Horwitz, Schapiro, Rapoport, Ungerleider, et al., 1992). The *what pathway* (or ventral stream) passes from the visual cortex to the inferotemporal cortex (IT), where information about the object becomes available for access by "higher" cognitive centers. The *how* (or *where*) *pathway* (the dorsal stream) channels the visual information from visual cortex to parietal lobe. It is this pathway that extracts parameters that are critical for physically interacting with the object.

One patient studied by (Goodale, et al., 1994b), referred to as D.F., suffered from a bilateral lesion of the ventrolateral occipital region (affecting the what pathway). Although she was capable of grasping objects just as normals would (preshaping her hand during the reach and placing her fingers on the objects in locations that facilitated a stable grasp), she was unable to discriminate between objects. In a second task, D.F. was repeatedly presented with a rectangular slot placed at various orientations (Goodale, Jakobson, Milner, Perrett, Benson, & Hietanen, 1994a). Although she was able to place a card within the slot, she could not indicate with her hand the orientation of the slot (the indicated orientation did not correlate with the actual orientation of the slot).

A second patient, R.V., suffered from a lesion in the occipitoparietal cortex, affecting the how pathway (Goodale, et al., 1994b). This patient showed no deficit in her ability to discriminate complex objects, but was unable to shape her hand appropriately in order to establish a stable grasp on the object. (Jeannerod, Decety, & Michel, 1994) reported a similar case, in which the patient had suffered a bilateral posterior parietal lesion. This individual, referred to as A.T., had apparently lost the ability to translate object size into hand aperture size. Consequently, she had developed a grasping strategy in which she preshaped during the reach movement by opening her hand to its greatest extent, and then clumsily enclosing her hand around the object once the arm had brought it within reach.

These observations suggest that these two visual systems are very distinct in the type of information that they compute and that functionally they solve a separable class of problems. However, it is a significant oversimplification to believe that these two pathways do not at all interact (Horwitz, Grady, Haxby, Schapiro, Rapoport, Ungerleider, et al., 1992). When D.F. is presented with a T-shaped object and asked to insert it into a T-shaped hole, she is able to accomplish the task on about half the trials (Goodale, et al., 1994a). On most of the remaining trials, the orientation that she chose was approximately 90^{o} from the correct orientation. Although the dorsal system has enough information to be able to align a piece of the object (e.g. the bar of the T) relative to one of the two slits that make up the T-shaped hole, it is the ventral system that provides the correct matching of the corresponding components.

An interesting interaction between the dorsal and ventral systems (or lack thereof) has also been observed in A.T. by Jeannerod (Jeannerod, et al., 1994). The most significant deficit in the grasping task is seen when the object being grasped is a neutral one - in other words, an object that is not part of A.T.'s everyday experience. When the object is one that is more familiar, the deficit is not as pronounced. In these cases, A. T. typically preshaped as one would normally.

In addition to the simple how/what distinction of visual processing, Jeannerod has gone further to partition the dorsal system into two distinct pathways (Jeannerod, 1994). One of these pathways is responsible for the representation of object shape and size, which is then used in the programming of the grasp. The other pathway captures the location of the object in space, providing information that can be used in selecting a goal position for the arm. This distinction explains D.F.'s ability to reach accurately towards an object while not being able to shape her hand appropriately.

Interaction of the How/What Pathways

The experiments performed with D.F. and A.T. have demonstrated that both the how and what pathways can individually compute grasp parameters for certain classes of situations. But how do these two pathways interact during grasping in a normal primate? One hint of this interaction is provided by experiments performed with a hemi-field blindsight patient (Toni, 1994).

Blindsight is a condition in which a lesion has occurred within the primary visual cortex, effectively removing it from the visual processing circuit (King & Cowey, 1992). Because it is the primary visual cortex that feeds many of the other visual processing areas, the visual deficit is very significant -- to the extent that subjects report that they are unable to see anything (Weiskrantz, Warrington, Sanders, & Marshall, 1974). However, it has been observed that subjects are still able to perform some visual tasks (most involving high-contrast or movement situations), although the patients are not consciously aware that they are using vision to perform the tasks (and in many cases have to be asked to "guess" at the answer). These observations are explained by the fact that the retina has an evolutionarily older pathway to the superior colliculus. The superior colliculus, in turn, provides some very limited visual information to the higher centers via the pulvinar (Goodale, 1993).

In the experiment of Toni, the hemi-field blindsight patient (only one hemisphere of the visual cortex had been damaged) is presented with a series of objects which must be grasped (bright objects against a dark background). With his "good eye" covered, the patient reported that he could not see the object presented to him. However, when asked to reach out and grasp the object, he was able to perform the task with reasonable accuracy. Moreover, just as with normals, the maximum grasp aperture (distance between the tip of the thumb and of the index finger) that was achieved during the reaching movement correlated with the size of the object. This implies that some visual information, via the superior colliculus, is available for aiding in the selection of grasp parameters.

In contrast, when the patient was asked to pantomime or verbally report the size of the object, the indicated size did not correlate with the actual size of the object. Furthermore, when asked to verbally report the size of the object during a reaching movement to the object, not only was the reported size not correct, the preshape was also incorrect. There was, however, a very strong correlation between the reported object size and the size of hand aperture during preshape.

What we can conclude from this set of experiments is that:

1. The visual information via the superior colliculus is not available to the what system.
2. This information is available to the how system.

3. When the what system is queried for an analysis of the visual scene and does not have any useful inputs, not only does it randomly select an answer, but it imposes this answer upon the how system.

Adaptation of the Visual/Motor Mapping

In a recent experiment, Gentilucci examined the potential plasticity in the mapping from visual input into grasp parameters in human subjects (Gentilucci, Daprati, Toni, Chieffi, & Saetti, 1995). The subject is seated in front of a horizontally placed mirror (Figure 4.1). An object that is to be grasped by the subject is placed beneath the mirror (this object is referred to as the *real object*). Due to the placement of the mirror, the object and the subject's hand and arm are not visible. A second object (the *virtual object*) is placed such that its image in the mirror is superimposed on the location of the real object. The result is that the subject believes that she is looking at the real object, but that during grasping is just not able to see her hand or arm.

During the first block of trials, the virtual and real objects are identical. During each trial, the subject responds to a go signal by reaching out and grasping the real object using a precision grasp; the kinematics of movement are measured during this process by tracking the 3-dimensional position of the wrist, and the tips of the index finger and thumb. As has been observed in earlier experiments (MacKenzie & Iberall, 1994; Paulignan, Jeannerod, MacKenzie, & Marteniuk, 1991), the maximum aperture achieved during the preshape phase of movement (the distance between the tip of the thumb and of the index finger) relates directly to the size of the object being grasped.

For the second block of trials, the virtual object is the same as in the first block of trials, but the size of the real object is changed by a small amount (+/- 10-12% in diameter). After completion of the second block of trials (12), Gentilucci observes the following:

1. The subject is not consciously aware that there is a difference between what is being seen and the object being grasped.

2. During the course of the second block of trials, the subject adapts the movement such that at the end of the block, the kinematics are as if the subject was both looking at and grasping the real object (i.e., the maximum aperture corresponds best to the size of the real object, not the virtual object).
3. When presented with the virtual object and asked to pantomime its actual size, the size indicated by the subject correlates with that of the real object, not that of the virtual object.

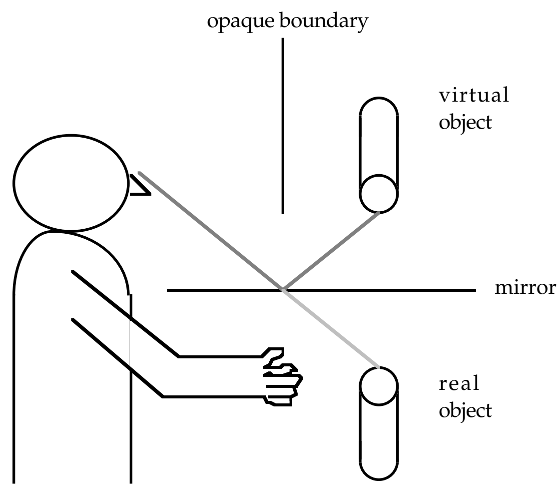

Figure 4.1: Experimental setup for the Gentilucci grasp adaptation task. The position of the virtual object is such that its image in the mirror corresponds to the location of the real object. An opaque boundary prevents direct viewing of the virtual object.

What observation #2 tells us is that there is a relatively quick adaptation of the mapping from the visual representation of the object to the parameters of the grasp that is subsequently executed. So, where is this adaptation taking place? The pantomiming task (observation #3) involves conscious access to the size of the object (note that D.F. also has difficulty with pantomiming the size of objects (Goodale, et al., 1994b)). Because this task is also affected by the adaptation that takes place during the second block of trials, we can conclude that at the very least, changes that are occurring involve the inferotemporal system (the what pathway) or a system that makes use of the inferotemporal object representation.

High-Level Model of Visual Processing

A high-level model summarizing the key experimental observations presented thus far is depicted in Figure 4.2. The primary visual pathway from the retina forks at the visual cortex

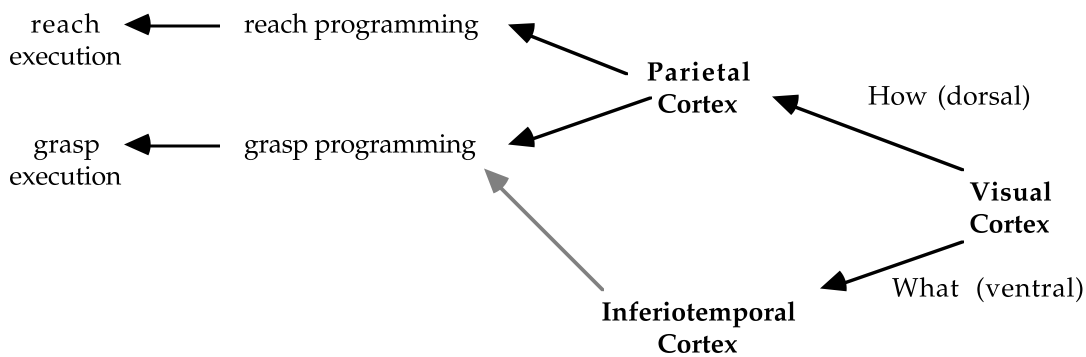

Figure 4.2: A high-level model of cortical visual processing and its support of the reach/grasp motor program.

into two functionally distinct pathways. The *what pathway*, leading to the inferotemporal cortex (IT), extracts the identity of the attended object, which in turn is associated with the object's properties (e.g., name, color and size).

The *how pathway* (Parietal Cortex) computes a parametric representation of the object being attended. In other words, motor-related descriptors are extracted that code information including the shape of the object, its size, and its location and orientation in space (Goodale, 1993). By *motor-related descriptors*, we mean that any object that is attended can be represented to some approximation with the descriptors, whether the object is a familiar one or not. In addition, two objects that share a common shape and size (e.g., a stick and a pen) will be coded in essentially the same way. In contrast, because these two objects are actually very different entities (in this case in color, texture and function), IT will utilize dissimilar representations.

Both the dorsal and ventral systems contribute to the selection of the grasp to be executed, as well as the grasp's associated parameters (in the figure these decisions are made in the *grasp programming* module). Here, *grasp programming* is distinguished from *grasp execution* in that it is a static representation of the grasp that is about to be executed. The grasp execution module is responsible for expanding this description into a temporal sequence of movements.

Because a single object is coded in two different ways in the dorsal and ventral systems, they contribute complementary information to the grasp programming process. The dorsal pathway provides a set of grasps that are specific to the shape and size of the object, which will often work for many situations. The ventral system provides context-dependent biasing of the grasp type (for selection purposes), and may also tune the grasp description (type of grasp and the associated parameters) for very specific situations that rely on context not carried by the shape and size of the object.

In the case of a parietal lesion (as with A.T.), the ventral system is the primary source of information for making the decision about the grasp (disregarding the sub-cortical contribution). When the presented object is a familiar one, a representation of this object exists within IT, from which an approximate grasp and accompanying parameters can be derived. However, when the object is one that is unfamiliar, its presentation does not activate IT, leaving the grasp programming module with no information. In human patients (specifically A.T.) this forces the adoption of other strategies for grasp programming.

Cortical Mechanisms for Object and Grasp Representation

In this section, we turn from human behavioral data to neurophysiological experiments performed in monkey. From this data, a more detailed model is derived of the cortical regions involved in seeing, reaching, and grasping. We first focus on the brain regions involved in the visual processing (Parietal Cortex and Inferotemporal Cortex), then turn to representation of grasps (F5), and finally discuss the mapping of object to grasp (AIP). These discussions are then summarized with an updated model.

The Dorsal System in Monkey

As has already been discussed, the dorsal system is made up of two functionally distinct sub-pathways. One stream carries object-centered information (shape and size), while the other represents the object's location and orientation in space. In monkey, distinct brain regions have been identified which are responsible for each of these classes of information.

The location of target objects is represented in the *Ventral Intraparietal area* (VIP) (Colby, Duhamel, & Goldberg, 1993). The coordinate frame appears to be a body-centered one, in that neurons are not modulated by the monkey's head or eye position. Locations in space are represented using a broadly-tuned population code. In this coding scheme, an individual neuron fires maximally for a small region of space (termed the *preferred region*). This activity drops off as the target position moves away from the preferred region. Although the individual neuron is active for a rather large region of space, the activity pattern across a large set of neurons can code positions rather accurately.

A neighboring region, the *Posterior Intraparietal area* (PIP), has been identified by Sakata as coding object-centered information (Sakata, 1994). Here, non-overlapping populations of cells have been shown to fire in response to different shapes that are presented to the monkey (common shapes used in the experiments included cylinders, plates, spheres and tori). Within a single population of cells, some showed specificity towards different parameter ranges (e.g., a *cylinder cell* might be responsive only to cylinders of width 10-20 mm). More generally, sub-populations of shape cells code the relevant parameters using a population-like coding scheme. Further details of this object representation scheme will be discussed in Chapter 5.

Note that Sakata does not claim (nor is it claimed here) that the monkey brain represents objects in terms of geometric primitives. Rather, these primitives act as a convenient partitioning of the objects into classes for testing of the monkey and for discussing what has been observed.

Neurons in PIP that code for the 3-D orientation of objects have also been identified (Sakata, 1994). A population code is used here as well -- with an individual neuron firing maximally for *preferred orientations*.

The Monkey Ventral System

(Logothetis & Pauls, 1995; Logothetis, Pauls, & Poggio, 1995; Tanaka, 1993) have examined cell responses in IT as a function of the object configuration that is presented to the monkey. In general, individual cells seem to be responsive to complex combinations of object properties. The cell responses were generally independent of the object orientation in the coronal plane, but did show specificity towards certain orientations in depth. In addition, many cells were not affected by the retinal size or position of the object's image.

A number of related experiments have looked at the population response to different types of visual stimuli. One key finding is that individual objects (or faces) are represented by sparse populations of neurons (Rolls & Tovee, 1995; Young & Yamane, 1992). This fact, combined with the fact that individual cells respond to complex combinations of features implies that two distinct objects are very likely to be coded by non-overlapping (or only slightly overlapping) populations of neurons.

Finally, several authors have also noted that cells within IT are capable of rapidly becoming responsive (on the order of minutes) to novel objects that are relevant to performing the task at hand (Miyashita, Date, & Okuno, 1993; Rolls, 1995).

High-Level Representation of the Grasping Program

The inferior premotor cortex (composed of F4 and F5) plays an important role in the high-level representation of both the reach and grasp. We have already reviewed in detail the data related to this region (Chapter 2). Here, we will summarize the key computational points and show how this region interacts with the visual areas already discussed.

Neurons in one subregion of Rizzolatti's area F4 have been hypothesized to code the goal position of reaching movements (Fogassi, Gallese, Fadiga, Luppino, Matelli, & Rizzolatti, 1995). Activity of single cells are sensitive to the goal position of the movement, but not to the starting position of the arm nor to the position of the head or eyes relative to the body. The positional goals are represented using a population vector-type representation, with individual neurons being active for rather large areas of space. Although F4 receives a very significant projection from VIP, it is not at present clear whether the space coded by F4 can be described by a peripersonal coordinate system (as is the case in VIP) or a joint-centered coordinate system. Temporally, we see activity related to the phase of movement: preparatory (or set), movement, and hold of position.

Area F5 is involved in the control of distal movements, with the dorsal half relating to the control of hand movements (Rizzolatti, et al., 1988). The key properties of individual cells in this region are as follows:

1. Cells tend to be most responsive for specific grasp types (or oppositions) made by the monkey. Cells coding for grasps that are not being executed are sometime inhibited.
2. The largest number of cells are dedicated to grasps involving the thumb opposing the fingers (pad opposition or precision grasp); a moderate number of cells is involved in the coding of side opposition; and a few cells are active during power grasps.
3. Most cells are most responsive to a particular phase of movement. The primary identified phases are: set, extension, flexion, and hold. The largest number of cells are active during the extension and flexion phases of movement.

Transformation from a Visual Representation to a Grasp

Up to this point, we have discussed cells that code information about the properties of the object to be grasped (in PIP and IT), as well as cells that code for the grasp that interacts with the object (in F5). In this section, we implicate AIP as the intermediate area responsible for the transformation of object to grasp.

Anatomically, AIP receives significant projections from PIP (Sakata, 1994), and exchanges heavy recurrent connections with F5 (Jeannerod, Arbib, Rizzolatti, & Sakata, 1995). Neurophysiologically, as we have already seen in detail in Chapter 2, individual cells exhibit behavior that reflects both object-related and grasp-related properties (Taira, et al., 1990). The key properties of these cells are summarized as follows:

1. Many cells within this area are selective for the type of object that is presented. Most often, cells are actually selective to varying degrees for several objects. Typically, these objects share some common characteristics, and may also be grasped in the same way.

2. Many cells show significant modulation in correlation with the phase of movement. The identifiable phases include: the *key* (set phase), extension (preshape), flexion (enclose), hold, and return. Cells participate in varying degrees during different phases of the movement, but most are highly active during the extension and flexion phases of movement.

3. Once a cell becomes active in the process of executing a reach and grasp, in most cases it will remain active until the task is completed (the time at which the object is released).

4. A few cells respond when simply looking at the object (Sakata refers to these as *visual-related cells*), others are modulated only by the movement to manipulate the object (these *motor-related cells* are even active when the monkey is behaving in the dark). However, most cells fall somewhere between these two extremes.

5. Some cells demonstrate specificity toward the size of the object to be grasped.

6. A small number of cells show modulation based upon the object's position and/or orientation in space.

Expanded Model

An expanded version of the high-level model is presented in Figure 4.3. Comparing with the original model (see Figure 4.2), the strictly functional modules (reach/grasp programming/execution) have now been assigned to particular brain regions.

VIP codes target object position and orientation in space using a population code in a peripersonal coordinate system. This target position is passed to F4, which represents the arm goal position. The distinction that was initially drawn in Figure 4.2 between the *programming*

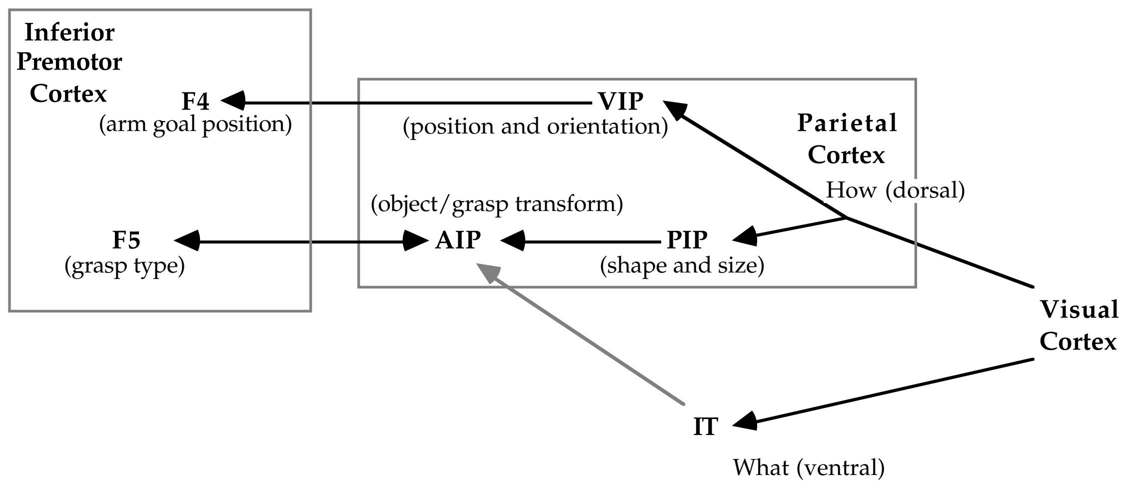

Figure 4.3: The expanded seeing/reaching/grasping model. The primary areas are IT (inferotemporal cortex), VIP (ventral intraparietal), PIP (posterior intraparietal), AIP (anterior intraparietal), and F4 and F5 (of the ventral premotor cortex).

and the *execution* of the reach has not gone away, but instead both have been absorbed into F4. Within F4, it is the *set cells* that are responsible for setting up the initial reach program, and then the *movement-* and *hold-related* cells that are responsible for executing the movement and maintaining the arm position, respectively.

As discussed above, PIP captures the shape and size of the object to be grasped. The details of this coding in the model will be discussed in Chapter 5.

AIP is hypothesized here as the first stage in the *grasp programming* process. As such, it is responsible for integrating object-related information from both the dorsal and ventral systems. The situation, however, is actually quite a bit more complicated in that AIP also receives grasp-related information from F5. Given that there are heavy recurrent projections between F5 and AIP, what are the differences in their behavior and hence their relative roles? The two key differences in behavior are:

1. Cells in F5 do not exhibit object-specific activity patterns (only grasp-specific behavior is observed), whereas we see both object- and grasp-specific activity in AIP.

2. Cells in F5 are active for only a small period of time during the execution of the movement. In AIP, once a cell becomes active (most are active by the beginning of the extension phase of movement, a few do not become active until the flexion phase), it remains active for the entire task, including the holding phase.

One interpretation of these differences (especially #2) is that AIP is serving as an *active memory* for the grasp that is about to be executed or that is being executed by F5. In other words, AIP is responsible for remembering the selected grasp, and F5 is responsible for unfolding this grasp in time during the execution of the movement. But why is it the case that F5 must send recurrent connections back to AIP? This would allow F5 to update the active memory with information about the grasp that is *actually executed*, which might differ in some ways from the grasp that was initially specified. Note here that we have deviated from the traditional neuroscience usage of the term *memory*, which is typically taken to mean that information is simply being held for some delay period so that it can be used later (e.g., Funahashi, Bruce, & Goldman-Rakic, 1993). Rather, we have taken a broader notion, in which the memory is continuously being updated as the result of other computations (in this case, as F5 monitors the progress of the grasp execution). This is similar to Dominey's *dynamic remapping*, in which motor afferent signals were used to update a model of the state of the eye plant as a saccade was being executed (Dominey, 1993). Here, the difference is that ascending sensory inputs can also be used to update the model state.

Execution of the Grasp

In this section, we will first review experimental data that focuses on several cortical regions that are involved in the execution of the grasp. We will then further develop the model to incorporate this data, and finally show a more detailed perspective on the interaction between AIP and F5.

Cortical Mechanisms of Low-Level Grasp Control

Feedback sensory information regarding the hand flows via the thalamus (ventral posterior lateral nucleus) through the primary somatosensory cortex (SI, which is composed of Brodman's areas 3a, 3b, 1, and 2). SI is a topologically organized strip that runs in parallel with the primary motor cortex, posterior to the central sulcus. Most cells in SI are responsive to only one modality: kinesthetic (areas 3a and 2) or tactile (areas 3b and 1) (Pons, Garraghty, & Mishkin, 1992). For each modality type, we see cells that reflect the basic receptor classes: slowly adapting (giving absolute information), and fast adapting (giving derivative information). The receptive field of a single unit in areas 3a and 3b tend to be restricted to small regions (on the hand); in areas 1 and 2, cells are responsive to larger regions of the hand.

The secondary somatosensory cortex (SII), receiving inputs from SI, provides a higher-level perspective of the current state of the hand. Many cells are responsive to both modalities and have receptive fields covering very large areas of the hand. These higher-order cells are capable of coding for very specific global hand configurations (Corkin, 1978). A lesion of SII in human results in a deficit in tactile recognition of objects, although the subjects are capable of describing the individual features of the object. In addition, a recent PET study examined the differences in pointing at an object versus grasping of the object (Grafton, Fagg, Woods, & Arbib, 1995). Subjects were instructed to grasp the object *very gently* and hold the position for a short period of time before releasing. One of the key findings of the study was that SII was more significantly active in the grasping case. Most likely this result is directly related to the fact that the subjects were instructed to very carefully grasp the objects.

The primary motor cortex (referred to as MI, Brodman's area 4, and F1) is one of the main sources of the corticospinal pathway (which supplies motor commands to the muscles) (Kandel, Schwartz, & Jessell, 1991). In the case of the hand, however, an even larger proportion of the corticospinal tract originates from MI. Microstimulation of the hand area of MI produces visible twitches in the fingers of the hand (Picard & Smith, 1992a; Picard & Smith, 1992b). An individual PTN MI cell typically projects to a set of muscles that are involved in controlling a single joint or several neighboring joints. We shall refer to this set of muscles as a *motor assembly* .

MI receives projections from many regions; here we are concerned with two sources. F5 sends motor commands to the hand via MI (Matelli, 1994). In our model, a group of F5 cells forms a coarse code for the grasp that is being executed. Therefore we imagine an F5 cell as selecting a set of motor assemblies that are relevant to implementing that grasp. We will return to this point in the next Chapter.

Sensory information from SI (and thalamus) can also activate motor cortex cells, implementing one level of reflex loops . Picard & Smith (1992a) examined the modulation of motor cortex activity by cutaneous and proprioceptive information during a precision grasping task. Two monkeys were taught to grasp an object with a lateral pinch (side opposition). For any particular trial, the object was chosen from one of three weights and one of four surface textures (oiled metal, smooth metal, and fine and coarse sandpaper). The monkey was allowed a view of the object before it was grasped. The grasped object was instrumented with a position sensor, as well as force sensors that measured horizontal grasping force and vertical lifting force.

Grip force was examined (during dynamic and static phases of movement) as a function of object type. For a particular surface type, grip force increased with weight of object. In all but one case, larger forces were applied to the more slippery objects. Most cells sensitive to weight changes increased in their activity, but a few showed a reduction in activity. Figure 4.4 compares the receptive field and motor output of a single cell that increases in activity as the object weight is increased. The cell was responsive to cutaneous stimuli on the lateral side of the index finger. However, stimulation of the cell resulted in abduction of the thumb.

Model of Grasp Control

The component of the model responsible for the execution of the grasp is shown in Figure 4.5. As has already been discussed, the grasp programs in F5 are represented by units that participate in a specific grasp and a particular phase (for purposes of the discussion we can assume that a single cell participates in exactly one grasp and during one phase). For phases in which movement is required (i.e. extension, flexion, hold, and release), the F5 cells drive the hand by activating very specific sets of primary motor cortex (MI) cells. This activation can either specify equilibrium positions for the joints of the hand (Katayama, 1993) or can configure reflex loops in which MI cells respond to inputs from SI (Picard & Smith, 1992b). More details of this implementation will be discussed in the following chapter.

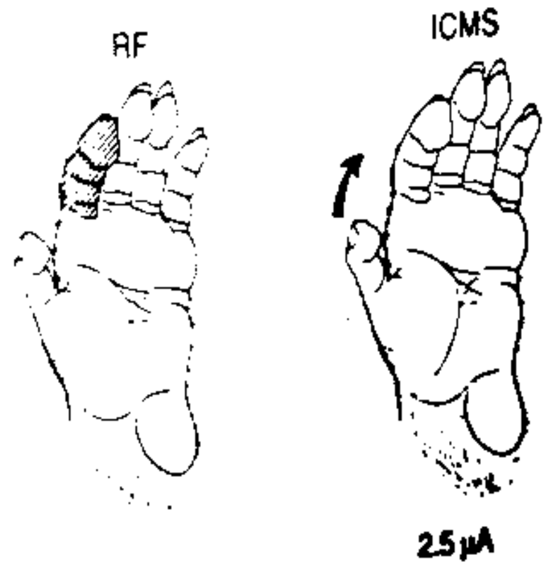

Figure 4.4: Motor cortex cell whose firing correlated with cutaneous stimulation of the lateral side of the index. Direct stimulation of the cell caused abduction of the thumb (Figure 4 of Picard & Smith, 1992a).

SII cells are responsible for detecting specific combinations of SI activity patterns (sensory hyperfeatures). These hyperfeatures are used in two ways:

1. Signaling of key events in the execution of the grasp, which can be used by F5 to trigger transitions in the phase of the program (e.g. transition from the flexion to the hold phase when contact with the object is detected).
2. Recognition of the object being grasped, which allows for refinement or reconfiguration of the original vision-based program.

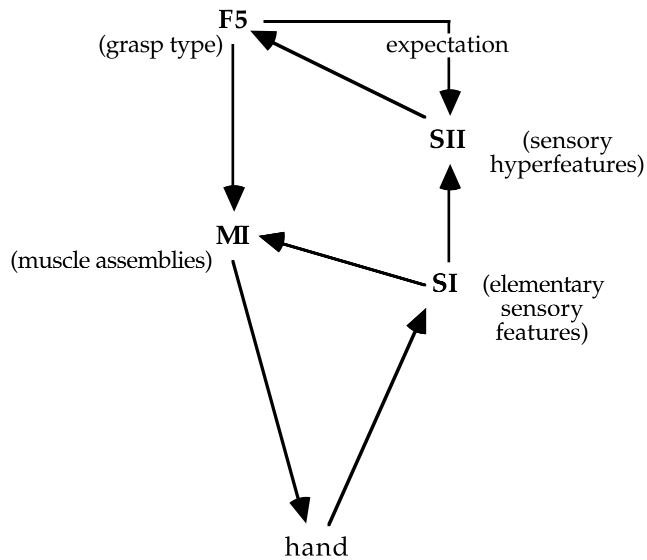

Figure 4.5: The cortical regions involved in the low-level control of the grasping motor program. The areas are F5 (inferior premotor cortex), MI (primary motor cortex), SI (primary somatosensory cortex), and SII (secondary somatosensory cortex).

In addition to the inputs received from SI, SII cells can also receive inputs from F5 units. These inputs express an *expectation* of coming sensory events (for example contact with a particular object), which allows the SII cells to more quickly and more significantly respond to the detection of the event. This expectation is also an important mechanism for filtering out incidental events which are not relevant to the current grasp (or phase) being executed.

Because expectation alone should not necessarily cause an SII cell to become active, a fundamentally different type of input is used. We refer to this input as a *priming signal* (Dakin, 1986). An individual SII cell must receive both a priming signal (from F5) and an *activating signal* (from SI) in order to fire. We will return to this distinction in Chapter 5.

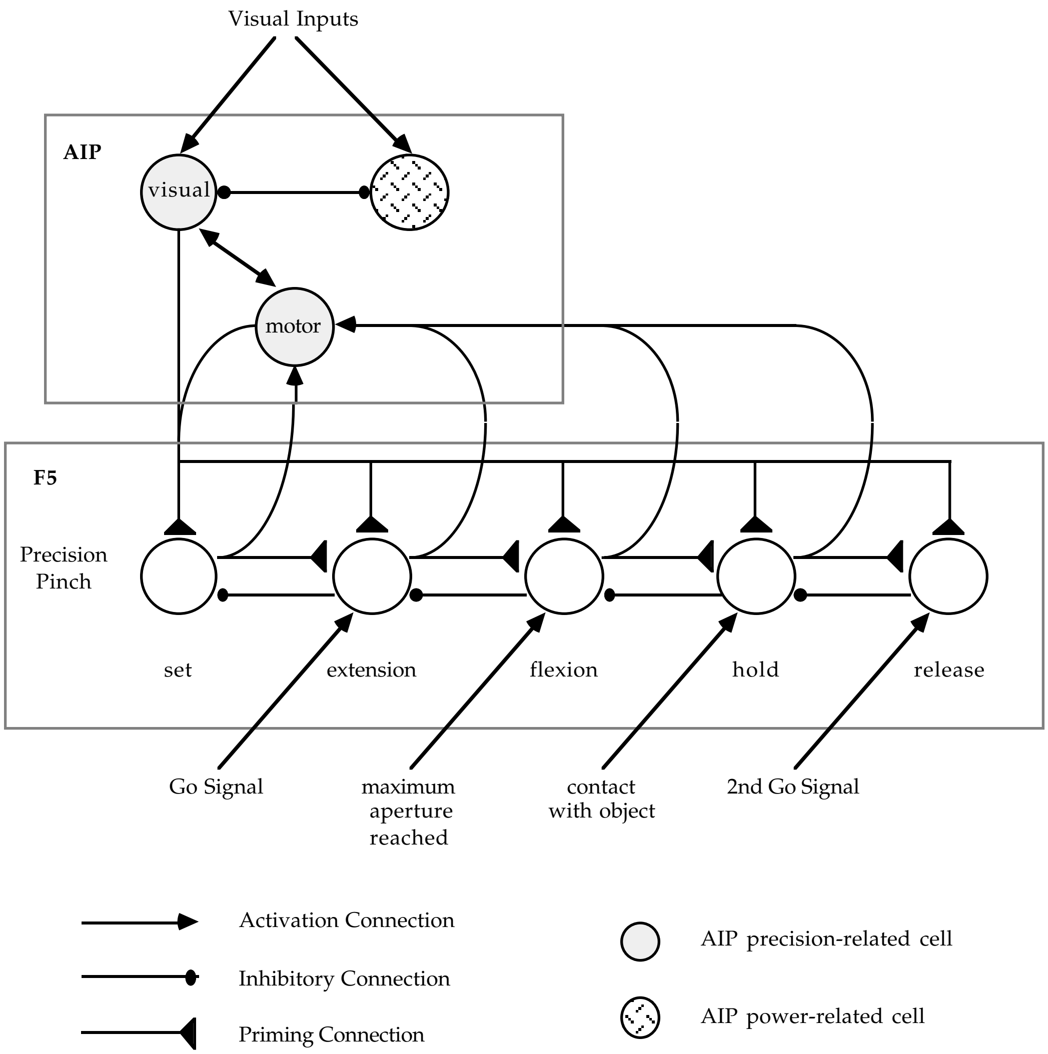

Figure 4.6: Interaction between AIP and F5 populations (circles) during execution of the preshape/grasp/hold program. Activity within F5 cascades from left to right as the program is executed. At each program phase, the state is reported back to the AIP motor-type population.

Interaction of AIP and F5 During Execution of the Grasp

Figure 4.6 demonstrates at a schematic level the interaction between populations of cells in AIP and in F5, with particular emphasis on the active memory and phasic activation mechanisms. Three AIP cells[1] are shown: a visual-related cell that recognizes objects that

1 Although we are using the term *cell* here, we are really referring to a population of cells.

require a precision pinch, a motor-related cell of the same type, and a visual-related cell that recognizes objects requiring power grasps. The five F5 units participate in a common program (in this case, a precision grasp), but each cell fires during a different phase of the program. We illustrate the dynamics that take place between AIP and F5 by walking through one full task (in this case the Sakata paradigm).

Initially some set of visual parameters that is specific to a particular object activates the visual-related cell in AIP (the cell that is precision-grasp related). A competitive process takes place within AIP, shunting the activity of cells that correspond to other grasps (in this case with a power-grasp related cell).

This activation *primes* a set of cells within F5 that correspond to the grasp to be used (in this case the precision grip). As a result, the F5 set-related cell becomes active, signaling a preparation for execution of the precision grasp. The set-related activity prepares the sub-program for execution by priming the cell representing the extension phase of the motor program.

When the trigger (Go) signal is given by the experimenter, the extension-related cell begins to fire, causing the preshape of the hand to begin (downward motor commands are not shown). At the same time, the activation of this cell forces the set-related cell to turn off (via the backward inhibitory connection), and the flexion-related cell is primed.

At the time that the fingers reach an extension appropriate for the size of the object to be grasped (as detected by the internal model of hand state or by an SII unit, neither of which are shown in the Figure), the flexion-related unit is activated, marking the beginning of the enclose phase of movement. In response to this activation, the extension-related cell is turned off and the hold-related cell is primed. The next phase transition occurs when a tactile stimulus is detected (by an SII unit); the hold-related neuron activates, shunting the flexion-related cell, and priming the release-related cell. The hold neuron continues to be active until the external secondary go signal is received. This signal causes the activation of the release neuron and the shunting of the hold neuron.

At each phase of the program, the corresponding F5 unit sends excitatory input to the motor related cell in AIP. The recurrent priming signal from this AIP cell serves as a active memory of the motor program that is currently being executed. This memory ensures that at any point in the execution of the grasping program, the F5 cells that correspond to the precision grasp (in this example) are activated during the next phase of movement, and not cells that participate in some other grasp. We will return to the details of these dynamics in Chapter 5.

As presented, the neural circuit in Figure 4.6 appears as a rather rigid structure. The neural implementation, however (as we will see in Chapter 5), generalizes along several dimensions. First of all, the "populations" of cells (depicted by circles in the Figure) do not necessarily fall into discrete sets of cells. Different grasps (in F5 and AIP) or phases (F5) are able to share cells to varying degrees. Second, we do not imagine the connections implementing the phasic behavior as being hardwired into F5. In the following Chapter, we present a more general structure in which the basal ganglia (BG) are implicated as representing the essence of the phasic connections. We imagine that in future models, the functional connectivity of the different populations could be reorganized so as to create arbitrary assemblages of component motor programs (Arbib, 1990) and/or sequences (Dominey, 1995). Third, although we treat each phase as an equal entity in the model, the extension and flexion phases of movement are probably better interpreted together as a *grasp phase* in the context of motor units that are selectable by BG. The within-phase modulation that implements the preshape and enclose are most likely handled by mechanisms other than the BG.

Remaining Computational Issues

It is very often the case that a set of biological experiments yields an incomplete view of the system being studied. As part of the model-building process, however, the goal is to construct a complete story as to how the system functions. As such, we are forced to rely on other sources of information to constrain the modeling. Possible sources of additional information come from *computational constraints*, in which we ask the question "given that our model must perform some task or exhibit some behavior, and that it has certain information available to it, how does it transform this information into a form such that it can successfully perform its task"?

Here, we outline several key computational issues that have not yet been addressed experimentally, and in the next section show how the model evolves in light of these constraints.

Object/Grasp Mapping is Many-to-Many

Any object with which a human or monkey will interact has many different ways that it can be grasped. However, to this date, the experimental paradigms that have looked at the mapping from object to grasp have only focused on situations in which a single grasp is appropriate for any particular object that is presented to the monkey. How can this more general transformation take place (i.e., from an object to a set of grasps)? And when we have a set of grasps that are appropriate, what information is used to select one to execute?

In the model, we address these questions by expanding our view of the roles of AIP and F5. However, before we do this, it is appropriate to first introduce a somewhat new perspective on what it is that AIP might be computing. As we have noted earlier, cell activity within AIP can exhibit specificity towards both grasps and objects. One way to interpret this partitioning is that AIP is computing the set of *affordances* for the object being attended. Here, we use the term *affordance* to mean that the system is highlighting the features of the object that are specifically relevant for physically interacting with that object. Note that we are deviating from Gibson's use of the term (Gibson, 1950) in that we do not assume that the environment is providing the affordances, but rather they are being computed by the two visual pathways.

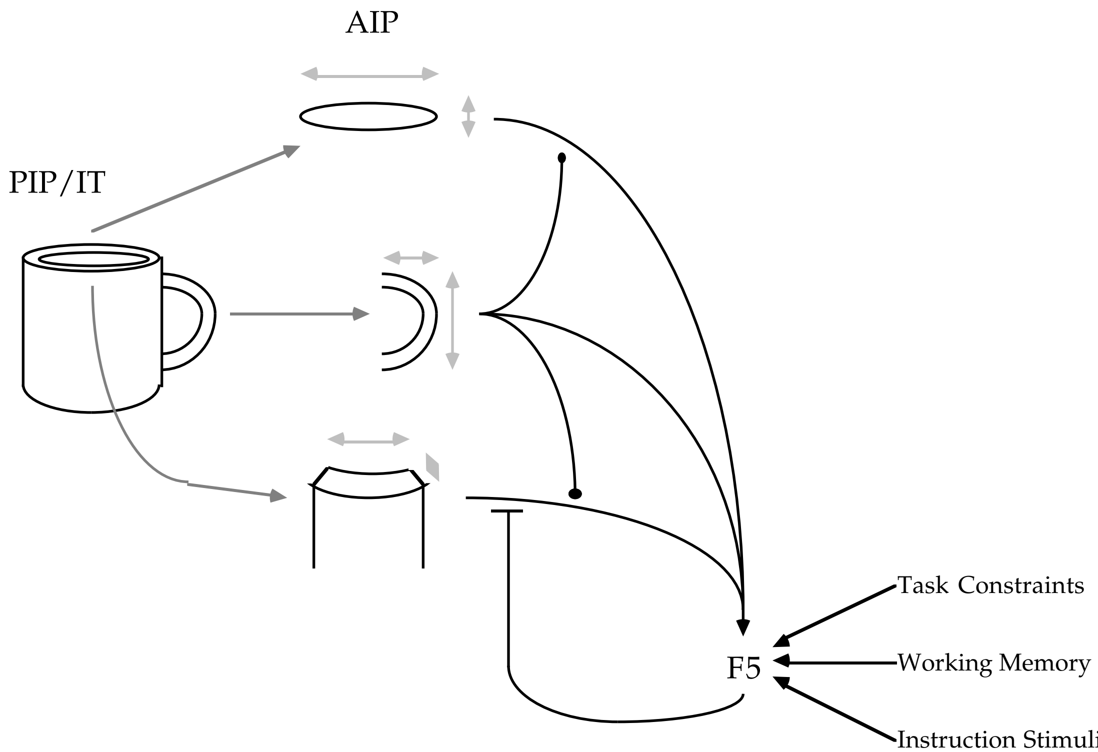

Figure 4.7: AIP computes the set of affordances for an attended object. These affordances highlight the features of the object that are relevant to physical interaction with it.

Figure 4.7 illustrates an example of computing the set of affordances for a mug. In this case, three different features are being highlighted. The circular rim of the mug affords a pad-opposition involving all four fingers (as in a grasp from the top of the mug). The handle of the mug affords a grasp that combines a palm opposition (e.g., index and middle against the palm) with a side opposition. Finally, a section of the rim of the mug affords a grasp in which the pads of the index and middle fingers oppose the thumb. Note that not only are the individual features highlighted, but the relevant parameters (e.g., the diameter of the rim) are also made explicit. Although this example has assumed one object feature is mapped to a single grasp, this is not generally the case. A cylinder, depending upon its size, can map to a precision pinch, a power grasp, or a lateral pinch.

When the mug is initially presented to AIP, based on inputs from both the dorsal and ventral streams, it computes the set of affordances. The corresponding set of grasps are passed to F5. As a function of task or abstract information, F5 selects one of the specified grasps. This decision is broadcast back to AIP, which shunts the other affordances, leaving only the affordance that corresponds to the selected grasp. During the execution of the grasp, the affordance represented by AIP is continually reinforced by inputs from the active grasp program in F5.

Modulation of the Reach Program by the Grasp Program

Introducing the possibility of a single object being mapped to multiple grasps brings an additional level of complexity--the wrist must be positioned such that the selected grasp can be established correctly (Arbib, Iberall, & Lyons, 1987; MacKenzie & Iberall, 1994). The implication here is that not only must the object position be taken into account (as represented in VIP), but once the F5/AIP system selects the grasp to execute, some form of modulatory signal must be combined with this representation in order to compute the final goal position of the wrist. One anatomical candidate for this modulatory signal is the projection from AIP to F4 (Matelli, 1994).

Parameter Coding

Experimental work on grasp coding in monkey cortex to date has focused primarily on the representational differences between grasp modes or objects of radically different shapes. However, in human behavioral experiments we can see subtle changes in the preshaping movement with small modifications to the size of the object (with the same grasp mode) (MacKenzie & Iberall, 1994; Paulignan, et al., 1991). How is it that these small differences in the object configuration and in the corresponding motor program are represented in PIP, AIP, and F5?

Sakata has observed that some AIP and PIP cells demonstrate sensitivity to the overall size of the object (Murata, Gallese, Kaseda, Kunimoto, & Sakata, 1993; Sakata, 1994). The fact that he observes cells that are both size-selective and size-independent indicates that within a population of cells that code for a particular affordance, a sub-population of these cells are responsible for capturing size of the object.

An important next question is: what is meant here by size? If we take as an example a cylinder, two parameters that describe the cylinder are diameter and length. In the experiments performed by Sakata these two parameters are scaled together, and never independently. If we could independently vary these parameters, how would this be reflected in PIP and AIP? Given that we have interpreted the cell behavior in PIP as simply representing the object and its parameters, we would expect to see these parameters reflected equally (approximately the same number of cells involved in each). However, if we interpret AIP as capturing only the information that is relevant to the physical interaction with the object, then a balanced representation may not be the case. Considering the cylinder, if the subject is performing a side opposition, then the diameter of the cylinder will be much more relevant to the execution of the movement than its length. Note that the length of the cylinder does not go completely unused--in the more general problem of which grasps are afforded by the object, this parameter could play a key role.

The Full Region-Level Model

Further Biological Background

The supplementary motor area (SMA) has been implicated in the planning and execution of complex movements. In human PET studies, we see a larger degree of activity in SMA during execution of a complex sequence of finger movements than for simple flexion of an individual finger (Roland, Larsen, Lassen, & Skinhøf, 1980). A number of studies have also demonstrated an effect in SMA related to mental imagery of complex tasks (Grafton, Arbib, Fadiga, & Rizzolatti, 1996).

In monkey, a unilateral lesion of the SMA has been shown to disrupt the monkey's ability to allocate his hands to different subtasks of a bimanual task (Brinkman, 1984). In addition, the SMA appears to be involved in the temporal organization of complex movements (Tanji, 1994). Based on cytoarchitectonic and microstimulation evidence, Luppino et. al argue that there are actually two distinct areas within what has been traditionally called SMA (Luppino, Matelli, Camarda, & Rizzolatti, 1993; Luppino, Matelli, & Camarda, 1991; Luppino, Matelli, & Rizzolatti, 1990). These two regions are referred to as SMA-proper (F3; the caudal region), and pre-SMA (F6). F3 is somatotopically organized and has heavy projections to the limb regions of F1, as well as direct projections to the limb-related portions of the spinal cord. F6 does not project to the spinal cord, and has only moderate projections to areas F3 and F2 (the dorsal premotor cortex) (Luppino, et al., 1990). However, there is a very heavy projection to area F5. Inputs into area F6 include VIP, and area 46 (principal sulcal region of the prefrontal cortex). Recording studies by Rizzolatti have observed neurons that become active when an object that the monkey is about to grasp moves from being out of reach into the peripersonal space of the monkey (Rizzolatti, 1990). The interpretation of this neural response is that this class of neuron is responsible for generating a *go signal* when it is appropriate for the monkey to begin a reaching movement.

Area 46 has been implicated as a working memory in tasks requiring information to be held during a delay period (Kandel, et al., 1991; Quintana & Fuster, 1993). This memory can participate in the learning of tasks involving complex sequences of movements (Dominey, 1995). Anatomically, this region projects to F6, and also exchanges connections with area F5 (Luppino, et al., 1990; Matelli, 1994). In human, (Decety, Perani, Jeannerod, Bettinardi, Tadary, Woods, et al., 1994) have recently discovered that area 46 is involved when a subject is asked to imagine herself grasping an object. In addition to this region, they observe activity in area 44 (possible F5 homologue), as well as a site along the intra-parietal sulcus (Grafton, et al., 1996).

Dorsal premotor cortex (F2) is thought to be responsible for the association of arbitrary stimuli (an IS) with the preparation of motor programs (Evarts, Shinoda, & Wise, 1984; Kurata & Wise, 1988; Mitz, Godshalk, & Wise, 1991; Wise & Mauritz, 1985). In a task in which a monkey must respond to the display of a pattern with a particular movement of a joystick, some neurons in F2 respond to the sensory-specific qualities of the input. However, many units respond in a way that is more related to the *motor set* that must be prepared in response to the stimulus. When a muscimol lesion in this region is induced, the monkey loses the ability to correctly make the arbitrary association (Kurata & Hoffman, 1994).

The Complete Model

The complete region-level view of the model is shown in Figure 4.8. This figure brings together our earlier models for visual processing and grasp execution, and incorporates the biological data discussed in the previous section. In the following Chapter, we describe the neural-level implementation of these circuits, focusing on PIP, AIP, F5, MI, SI, and SII.

The projection from AIP to F4 carries wrist position modulatory information that is to be combined with the object location information flowing from VIP. This modulatory information is specific to the single affordance/grasp pair that is selected by the AIP/F5 system.

Regions 46 and F6 are responsible for representing the high-level program. F6 first prepares the ventral premotor regions for execution of the coming reach and grasp by priming both F4 and F5. It then detects the *go signal* given by the experimenter and initiates execution of the program (generates the go signal shown in Figure 4.6). Initiation of the *release phase* of movement is also handled in this way.

Area 46, working in conjunction with F6, is responsible for supplying any task-dependent biases for the selection of the grasp in F5. This selection can be based upon the task

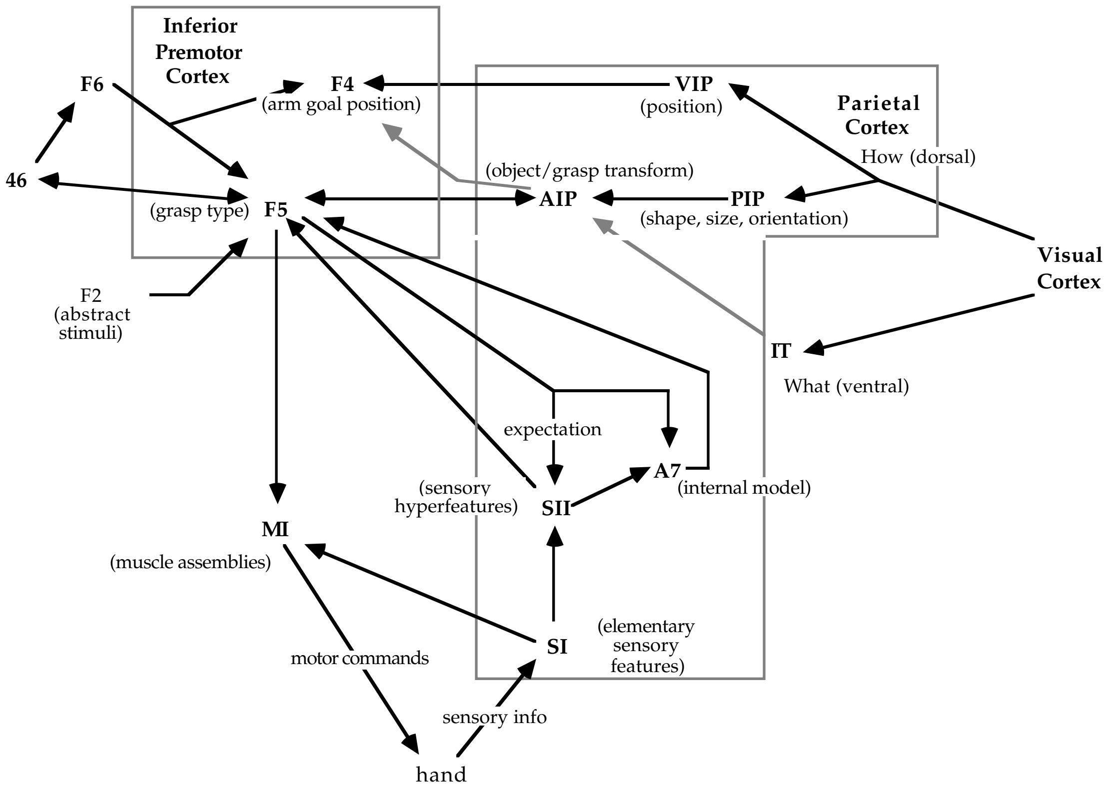

Figure 4.8: The complete model. The two key additions to the model are the projection from AIP to F4, and the introduction of regions 46, F6, F2, and area 7.

requirements (such as what is going to be done after the grasp), or based upon a working memory of a recently executed grasp. The biasing can be done at the level of the class of grasp (e.g. power versus precision), or can also include the parameters of the grasp (e.g. width of the aperture).

Progression of Activity During Task

In this section, we chart the flow of information through the model as a variant of the Sakata task is performed. The steps are similar to those presented at the beginning of this chapter.

Step 1: The trial begins with F6 preparing to receive the *ready signal*.

Step 2: The ready signal is presented and the target object becomes visible. F6 responds by priming F4 and F5. Area 46 extracts any task-specific information that is available and computes a grasp bias which is passed to F5.

The visual information flows from the visual cortex to the inferotemporal and parietal cortices. IT identifies the object (if it is a familiar one). VIP computes the object's location in peripersonal space. PIP extracts the parametric description of the object (shape and size). The resulting activity patterns in PIP and IT activate a set of affordances in AIP.

The active affordances in AIP excite the set of corresponding grasps in F5 (only the set cells in F5 are activated). Within F5, the active grasps compete through a winner-take-all mechanism (Didday, 1976; Fagg & Arbib, 1992), which incorporates any biases that might be received from area 46. The final remaining grasp is reported back to AIP, where the other affordances are shunted. The selected affordance sends wrist modulatory information to F4, where it is combined with the location of the object from VIP.

Step 3: The flash of the LED (the go signal) is detected by F6 and then passed on to F4 and F5. Within F5, the set-related activity gives way to activation of the extension-related cells (we assume that the F4 execution mechanism is similar to that of F5). These cells initiate the preshape movement by activating the appropriate set of motor assemblies within MI. In addition, event detectors in SII are primed that are responsible for sensing the point at which the maximum aperture is reached. As execution of the extension progresses, the internal model of hand state is updated from the descending motor commands. In parallel with SII (but without the sensory delay), the internal model estimates the point at which the hand has reached maximum aperture (we implicate area 7 as implementing the internal model).

When this event is detected, the flexion-related cells in F5 are activated. In turn, the extension cells are shunted, the flexion movement is initiated within MI, and the SII cells responsible for detecting contact with the object are primed. The MI motor assemblies that are activated involve not only an absolute movement (to the expected width of the object), but also involve reflex loops that will react (at a joint level) to sensed forces.

Step 4: Contact with the object is established by all fingers involved in the grasp. This event is detected by SII. In F5, we see a phase transition from flexion to hold. The hold cells in F5 activate the MI motor assemblies that are involved in maintaining a stable grasp. In addition, these cells prepare for the release event by priming the F5 release cells.

If there is a miscalibration between the expected and actual global hand states (e.g. in the case of an object with a width that is different than what was assessed by the visual system), then this is detected in the model by SII. The result is that a different set of SII units will become active than the set that was primed by the F5 flexion units. In F5, this means that the activated hold cells will correspond to the object that is actually grasped, and not to the set that would normally become active when there is not a perturbation. This change in grasp program is propagated to AIP, where we see a shift in the representation of the affordance.

Step 5: The reception of the secondary trigger stimulus is detected by F6, which forwards the trigger on to F4 and F5. F5 activates the motor assemblies that cause the hand to open just enough to release the target object. This event is detected by SII. F6 then shunts its priming signal to F5 and F4.

Summary

In this chapter, we have outlined the primary cortical regions that are involved in the seeing/reaching/grasping process and have assembled a high-level model of these processes. The key components of the model and their contributions are:

1. The dorsal visual stream (parietal cortex) extracts parametric information about the object being attended. This information provides the details of the object configuration that are used to finely tune the grasp plan. The representation is general in that both familiar and unfamiliar objects are represented equally.

2. The ventral system (inferotemporal cortex) extracts information that is used for conscious processing of the object. In the grasping task, this region contributes by biasing the grasp selection process towards grasps that are most appropriate for this specific object. The representation is a localist one, in which an individual object is coded by a sparse set of neurons, and different objects are coded by non-overlapping sets of neurons.

3. The anterior intraparietal area (AIP) is hypothesized as playing a dual role in the seeing/reaching/grasping process. First, it is responsible for integrating visual information from both the dorsal and ventral streams, and computing the set of *affordances* exhibited by the object. Second, as one of these affordances is selected and execution of the grasp begins, AIP behaves as an *active memory* of the one selected affordance and updates this memory to correspond to the grasp that is actually executed.

4. F5 is hypothesized as first being responsible for integrating task constraints with the set of grasps that are afforded by the attended object in order to select a single grasp. After selection of a single grasp, F5 unfolds this represented grasp in time to perform the execution. During execution, F5 monitors the progress and any changes to the program (due to unexpected events) are reported back to AIP so that the active memory can be updated.

In the following chapter, we further develop the neural-level details of the model.