Chapter 3

Visuomotor development

Sensorimotor coordination has been an active research topic for both neuroscience and artificial intelligence over the last decade. The integration of sensory information for movement guidance represents perhaps the most basic operation that a nervous (or artificial) system must solve (Churchland and Sejnowski, 1992). Despite exploring the same problem, studies carried out in these fields have arrived at different solutions. These differences arose mainly due to the divergent research goals.

Robotic models, rather than revealing how nervous systems attain sensorimotor control, demonstrate that many schemes are capable of implementing sensorimotor coordination and focus on the accuracy of controlled behavior, often without considering the biological plausibility of the resulting implementation (Kalaska, 1995). Nevertheless, their capabilities are still behind the adaptive motor abilities of biological organisms. Conversely, significant progress has been made in neuroscience towards understanding the biology of sensorimotor transformation in terms of integrative mechanisms that function at the single cell level or in small–networks of neurons (for a review see Kalaska et al., 1997; Snyder, 2000).

Several attempts were made recently to create an integrative framework that links neurophysiological and computational aspects of sensorimotor coordination. The aim is to provide a conceptual scheme that is readily implementable, to be used for the biological modeling of visually guided movements (Bullock et al., 1995; Burnod et al., 1999). The purpose of this chapter is to describe a general theoretical and modeling framework of sensorimotor development, inspired by these recent neurophysiological findings and computational theories.

Section 1 introduces the main research themes on biological sensorimotor coordination. First, a general description of the classical approach to solving the sensorimotor mapping problem is presented. Following this, we present a number of alternate hypotheses that have recently been proposed. Section 2 focuses upon the neurobiological mechanisms underpinning visuomotor mapping for arm reaching. Neurophysiological evidence is presented for the hypothesis of progressive match of visual to motor information, through intermediate spatial representations and supported by multimodal combinatorial properties of neurons. Finally, Section 3 reviews a number of biologically inspired models of visuomotor coordination development.

3.1 Sensorimotor coordination in biological systems

The general understanding of the sensorimotor coordination problem is that it represents a mapping of signals from various sensory modalities onto an appropriate set of efferent motor commands addressed to skeletal muscles or robotic actuators (Massone, 1995). A number of different aspects contribute to the difficulty of this task. First, investigations of the neural circuits involved in this process in the primate brain are faced with significant issues of complexity. This is because, it is not obvious how the perceptual areas are connected with the motor regions for sensory information transfer (Glickstein, 2000; Kandel et al., 2000). Second, in order to guide motor actions, the visual signals undergo a translation into motor outputs, which is commonly described as a non–linear coordinate systems transformation. Solving this mapping is again a non–trivial problem (Pellionisz and Llinas, 1980; Hildreth and Hollerbach, 1985; Jordan, 1996).

As a result, significant efforts in the last few decades have been devoted on one hand, to identifying the connections between sensory and motor areas in the brain (Tanne et al., 1995; Glickstein, 2000) and on the other hand, to developing various conceptual and modeling schemes that can solve the coordinate system transformation problem (see Kalaska et al., 1997; Snyder, 2000). These research directions are briefly discussed in the first part of this section, referred to, as the classical paradigm. Following this, we introduce three new theories, which refine the way researchers look at the visuomotor mapping problem.

3.1.1   The classical approach

Coordinate system transformation and the inverse kinematics problem

The concepts of reference frames and coordinate systems are used widely in the study of eye and limb movements. A *reference frame* is invoked when an experimental result is described: for example, describing the position of eyes relative to the head (a head-fixed frame of reference) (Soechting and Flanders, 1995). A *coordinate system* represents the set of axes fixed to the frame of reference, used to make the measurement (i.e., Cartesian coordinates). There are three main types of coordinate systems: *end-point coordinates* (e.g., retinal or Cartesian), *generalized coordinates*, that are independent variables describing the dynamics of a system (e.g., joint angles), and *actuator coordinates* dependent on how motor control is implemented (e.g., muscle forces) (Mussa-Ivaldi, 1995). For each coordinate system there are several possible reference frames: eye–centered, head–centered, limb–centered (Andersen et al., 2000).

The problem of visuomotor coordination was first described as a *coordinate system transformation* by Pellionisz and Llinas (1980). Given the fact that sensory input and the motor output vectors are represented in different coordinate systems, the transformation of sensory signals to motor commands can be described as matrix multiplication. And this represents the standard technique for solving the geometrical problem of going from one coordinate system to another (Pellionisz and Llinas, 1980).

The transformation of one class of coordinates to another is a nonlinear mapping. The translation from Cartesian coordinates to the joint angles required to move the arm to the target is referred as the *inverse kinematics problem*. The mapping from desired joint angles to muscle activities represents the *inverse dynamics problem*. The opposite process, which computes first the motor commands and translates them into hand movements, is referred as the *forward kinematics*. A biological/artificial system that has to perform the transformation from extrinsic to intrinsic coordinates must learn the spatial transformations (i.e., the inverse kinematics) and the muscles dynamics (i.e., inverse dynamics) (Ghez et al., 2000).

A common approach to sensorimotor learning in artificial systems is to learn first an inverse model, namely the transformation from desired movements to motor commands, which will further be used to control the end–effector (Massone, 1995; Jordan, 1996). There are various methods by which such an inverse model can be learned, mainly based on *error correction mechanisms*, which translate errors expressed in the Cartesian (visual) coordinate

system into the motor command or other coordinate system errors (for a review see Jordan, 1996; see also Section 3.3.2).

Nevertheless, it is an open question whether biological processes for inverse kinematics or inverse dynamics exist, or whether the nervous system effectively computes the spatial error between hand and target (Hildreth and Hollerbach, 1985). Despite the aptness of the coordinate system transformation concept as a description of neurophysiological processes, the question has been raised whether or not it has a biological relevance (Robinson, 1992; Kalaska, 1995; Burnod et al., 1999). Robinson (1992) argues that coordinate systems are a human invention for measuring spatial relationships and they are irrelevant for neural networks performing visuomotor transformations:

> Mathematical descriptions of what a system is trying to do are of little help to the neurophysiologist trying to understand how real neurons do it (Robinson, 1992, page 48).

Recently, a number of researchers acknowledged that even though neurophysiological data suggests that the brain does not literally implement this formalism, the *metaphor* of coordinate transformation does have heuristic value (Kalaska, 1995; Soechting and Flanders, 1995; Kalaska et al., 1997).

The connectivity problem

From the neurophysiological perspective, the sensory–to–motor transformation represents (partially) a problem of identifying the anatomical ways which connect the visual and motor areas. Since the primary sensory and motor cortical areas had been identified by 1900, most researchers have assumed that a series of cortico–cortical fibers must exist to link the perceptive and motor cortex for the sensory guidance of movement (see Glickstein, 2000). As Kandel and colleagues puts it, there exists a separation between motor and sensory systems and integration needs to be done:

> Purposeful action is possible only because the parts of the brain that control movement have access to the ongoing stream of sensory information in the brain. The integrative action of the nervous system depends therefore on the interaction between motor and sensory systems (Kandel et al., 2000, page 651).

The way visuomotor coupling is attained is particularly difficult to observe in the primates brain, due to the complexity of inter–neural circuits that stand between the sensory and motor neurons:
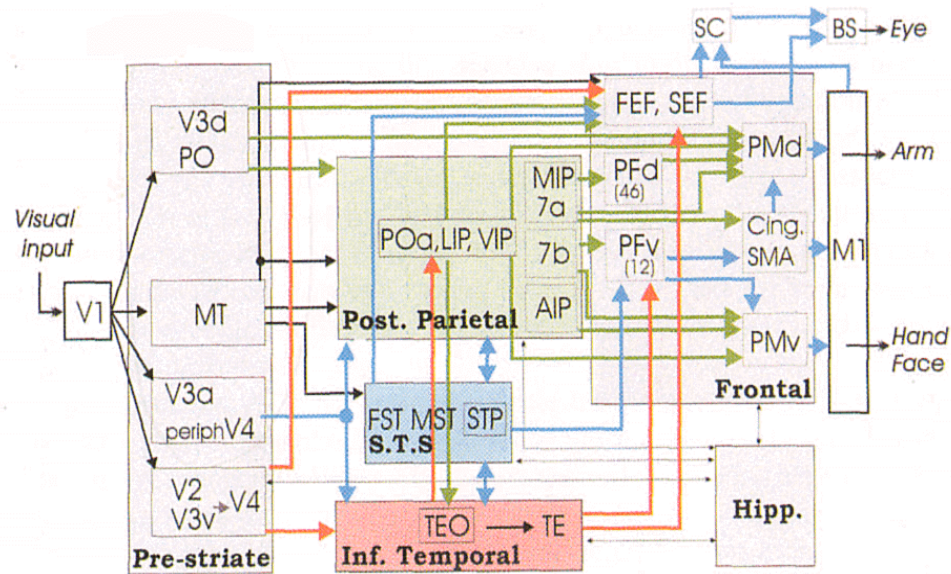
Figure 3.1: Cortical neural circuits allowing visual inputs from VI to be transformed into motor output in MI (adapted after Rossetti et al., 2000). The dorsal stream is represented in green, the ventral 'what' stream is shown in red and their possible interactions are drawn in blue. Legend: AIP: anterior parietal area; BS: brainstem; Cing: cingulate motor areas; d: dorsal; FEF: frontal eye field; Hipp: Hippocampus; LIP: lateral intraparietal area; M1: primary motor cortex; MIP: medial intraparietal area; MST: medial superior temporal area; MT: medio-temporal area; PM: premotor cortex ; SC: superior colliculus; SEF: supplementary eye field; SMA: supplementary motor area; STS: superior temporal sulcus; TE: temporal area; TEO: temporal-occipital area; v: ventral; V1: primary visual cortex; VIP: ventral parietal area.

> ...between the sensory receptors detecting signals and the motor neurons inner-
> vating the muscles are interneurons. Most of the answer to the problem of sen-
> sorimotor control resides in how these interneurons connect up to form circuits,
> and how these circuits are hooked up to produce behavior suited to the perceived
> situation (Churchland and Sejnowski, 1992, page 331).

Today, there is clear evidence that multiple-synaptic cortico-cortical pathways connect the primary visual and motor areas by way of the *dorsal stream* (Goodale and Milner, 1992; Tanne et al., 1995). Subcortical routes may also be involved in the visuomotor transformation process. An example is the route from visual to motor systems through the pons and cerebellum (Glickstein, 2000).

The dorsal pathway projecting from the primary visual cortex (V1) through the dorsal extrastriate visual area V3 and middle temporal area (V5/MT) to the superior temporal and parietal cortex is concerned with the perception of movement and spatial information (see

the green stream in Figure 3.1). Within the dorsal stream there are several pathways that are responsible with the regulation of reaching movements, grasping of the objects, or movements of the eyes as they follow a moving object (Andersen et al., 2000).

The connectivity problem is illustrated in Figure 3.1, which points at the complexity of the dorsal stream cortical circuits involved in the transformation of visual inputs into motor output (figure adapted from Rossetti et al., 2000). The dorsal stream (in green) is pictured along with the ventral 'what' stream (in red) and the substrate of their interactions is also shown. Within the dissociation paradigm, the study of Rossetti and colleagues explores the possible support for interaction between the vision and action systems.

To conclude, the biological theories outlined above helped to create a paradigm of a neatly partitioned brain into areas for perception and action. This in turn, influenced the way sensorimotor mapping problem was defined and addressed (i.e., as a transformation problem). It is believed in this thesis that the description of biological visuomotor development in terms of the inverse kinematics problem and coordinate system transformation cannot be taken literally, but rather as a *metaphor* of how the motor system processes information (Kalaska, 1995). It is also believed that understanding of how basic visuomotor coupling is implemented in the brain does not necessarily involve finding the connecting pathways from the primary visual cortex to the primary motor cortex.

The progress achieved in understanding the various functions of the neural populations placed along the dorsal stream and involved in movement control has changed the way researchers think about the sensory to motor transformation (Caminiti et al., 1996; Burnod et al., 1999). For instance, while reviewing neural information processing in those brain areas involved in the generation of sensory guided movements, researchers observed that there is not such a distinct line between sensory and motor processing in the brain (Goodale, 2000; Rossetti et al., 2000). In the remainder of this chapter, we focus upon the biological and computational hypotheses brought forward by some new ideas about visuomotor coordination development.

3.1.2   New theories of visuomotor coupling

This section presents new approaches of visuomotor mapping problem from three perspectives: neurobiological, computational, and evolutionary. The first of these frameworks is build upon neurophysiological data from studies of visually guided arm–movements in

monkeys. This research strand provided new insights into the organization of the cortical areas underlying reaching. Firstly, the cortical control of reaching is distributed along several areas in the parietal and frontal cortex (Johnson, 1992; Colby and Duhamel, 1996). Secondly, recruitment of neural populations from these areas for computation of motor commands involves both serial and parallel mechanisms (Caminiti et al., 1996; MacKay, 1996). Thirdly, the parieto–frontal network has a gradient architecture which favors the link of sensory and motor signals into a common hybrid reference frame (Johnson et al., 1996; Caminiti et al., 1998). Fourthly, the common frame is the eye–centered representation used in both ongoing and intended arm and eye movements (Andersen et al., 1997). Fifthly, neural cells from the parietal and frontal areas possess combinatorial properties by which they integrate information coming from various sources (Wise et al., 1997; Caminiti et al., 1998).

These findings have led to the reconsideration of the nature of visuo–to–motor transformation:

- The computational demand for reaching is met by operations that align distributed sensory and motor representations through unsupervised means. The nonlinear mapping is performed gradually, sustained by the combinatorial properties of individual neurons and the gradient architecture of the parieto–frontal network (see Section 3.2).

Another evolving research stream focuses on the use of computational modeling in an attempt to understand the mechanisms of adaptive behavior in autonomous agents (Arbib, 1987; Cliff, 1990). Research carried out on understanding simpler animals behavior indicated that all sensorimotor systems interact with their immediate surroundings by forming a closed loop with the environment (Cliff, 1995). It was suggested that:

- Sensorimotor pathways generating adaptive behavior might not be so precisely clustered into representation–transforming modules. Hence, one might try to model and understand *entire* sensorimotor pathways that are complete sequences of neural processing from the sensory input to the motor behavior (presented in Section 3.1.2).

Finally, whilst studying the interplay between eye and hand movement during reaching or manipulation of objects, researchers have realized that an essential collaboration may exist between oculomotor and limb motor control systems (Engel et al., 2000; Soechting et al., 2001). These behavioral and electrophysiological findings on eye–hand coordination suggested the hypothesis:

- Hand and eye movements are subject to similar control mechanisms and gaze provides the signal for targeted limb motion (presented in Section 3.1.2).

In a summary, the developing ideas outlined above promise to advance our knowledge of the biological and computational mechanisms of visuomotor transformation. They are rapidly maturing and we may soon witness the emergence of a new paradigm in biological visually guided motor control. The work presented in this thesis was partially motivated by the attempt to bring modeling support to these somewhat radical proposals. In the following, the latter two theories will be briefly described in subsections below, leaving the neurobiological mechanisms to be described in a dedicated section (Section 3.2).

The unitary nature of the sensorimotor cycle

One might assume, that being "probably the most basic operation a nervous system evolved to solve" (Churchland and Sejnowski, 1992) the sensorimotor coupling is implemented by some basic, primitive mechanisms on which sophisticated structures, like a primates brain, have built sensory guidance of movement (Goodale, 1996; see also the modeling of the entire sensorimotor pathways in computational neuroethology, Cliff, 1995).

The concept of a unitary sensorimotor cycle as a motor primitive for the generation of adaptive behavior in animals (and humans) is not recent. For a long time in biology, *the reflex arc* was assumed to play a central role in the production of complex sequences of movements, which were understood to be formed by the combination of simple reflexes (Kandel et al., 2000). More recently, computational neuroscience scholars in the search of primitives for sensorimotor development have studied and emphasized the importance of primitive and postural reflexes for infant development of motor control (Kuniyoshi and Berthouze, 1998; Metta et al., 1999). Metta and colleagues (1999) studied the issue of sensorimotor development within an artificial system and suggested that simple initial behaviors, such as motor reflexes and sensory-triggered motions, can be seen as computational *building blocks*. That is, they can guide the learning of more sophisticate behaviors and act as a bootstrap procedure for the whole system (see Section 3.3.1).

Further evidence for the unitary nature of sensorimotor pathways came from experimental and modeling studies of visually guided behaviors in less intelligent animals (such as amphibians and arthropods) (Arbib, 1987; Liaw and Arbib, 1993) and flies (Franceschini et al., 1992). In simpler nervous systems it is more evident that perceptive capabilities of an

organism have evolved in close relation with the achievement of specific motor goals. As Goodale puts it:

> Vision did not evolve to enable organisms to perceive. It evolved to provide distal control of their movements (Goodale 2000, page 365).

Neurophysiological data from simple vertebrate systems shows that different classes of visuomotor behavior rely on independent neural circuits specialized for generating an adaptive behavior to a certain sensory stimulus. For instance, in amphibians different circuits are involved for visually guided prey–catching and visually guided obstacle avoidance. Similarly, in rodents different subcortical pathways exist for control of orientation to food and novel stimuli and for the avoidance of obstacles during locomotion (Goodale, 1996). These observations suggest once more, that in a simple vertebrate's nervous system, vision and action are not so clearly partitioned in representation–action modules.

However, the complexity of mammals' lives has demanded more flexible organization of the circuitry, than that observed in simple vertebrates. According to Goodale (1996), this has been achieved by the development in the mammals' brain of a special system for perception, and particularly by the sophistication of the dorsal stream responsible for the perception–action loop. The ancient subcortical visumotor circuits that are shared by the monkey's brain with simpler vertebrates have become modulated by the more recently evolved layer of cortical control. This layer has evolved to make more adaptive motor behavior possible. The role of the inter–neural circuits in diversifying the motor behavior and allowing more complex response patterns has been explored also by modeling studies of the visuomotor pathways in the lamprey (Churchland and Sejnowski, 1992).

The classical view, as outlined in Section 3.1.1, defines the visuomotor mapping problem as the task of finding the path between the sensory and motor systems which implements the coordinate system transformation. The lesson from neuroethology is that perception–for–action is a crucial adaptive operation, which is implemented in less sophisticated nervous systems as unitary sensorimotor pathways (circuits). In this thesis, we consider an incremental approach to system construction. From this view, we believe that the development of new specialized modules or the increased sophistication of existing circuits preserves the unity of the perception–action cycle as a built–in, inherited mechanism.

Common computational programs for eye–hand movements

Within the visuo–motor mapping problem, special attention has always been given to eye–hand coordination. That is, because of its fundamental relevance to the organization of human motor skills and to the understanding of the perception–action relationship (Caminiti et al., 1992). Within this field, much consideration has been given to the possible interactions between the systems controlling the eye and arm. For a long time, the common assumption was that eye and arm movements may have little in common due to the fact that the eye is comparatively simple and predictable mechanical object, in contrast with the complexity of the arm system (Hildreth and Hollerbach, 1985). The traditional view of eye–hand coordination was that the central nervous system uses visual information (retinal or gaze angles) to build up representations of the environment and to guide limb movement (Kandel et al., 2000).

Recently, it was shown that for eye–hand coordination during manipulatory tasks, the gaze is well ahead of the hand and the object. This suggests that gaze supports hand movement planning (Johansson et al., 2001). Furthermore, Engel et al. (2000) demonstrated a similarity in the response of eye and manual tracking to a change in the direction of target motion. As a result, they proposed that hand and eye movements are subject to common control mechanisms and that gaze (i.e., extraretinal information) provides the signal for targeted limb motion. This hypothesis is also supported by neurophysiological studies which reveal that arm movements are coded and updated in eye-centered coordinates (Andersen et al., 2000; Snyder, 2000; see also Section 3.2).

These findings gave rise to a computational hypothesis about the way gaze signal might be 're-used' for arm movement control (Scassellati, 1998; Reilly and Marian, 2002). It is known that, in certain species, the auditory or visual signals produce an alignment of the head to the source of the signal. As a result, the motor head map is used to generate an appropriate motor behavior (Lazzaro et al., 1993; Kolesnik and Streich, 2002). Inspired by this idea, models of sensorimotor transformation have been built to simulate the sensory guided behavior of amphibians or flies (Franceschini et al., 1992; Liaw and Arbib, 1993).

The new hypothesis is that such a process might also operate in the case of motor control in primates, and to generate the neural program for arm movement the system might make use of the eye motor program (Metta et al., 1999; Reilly and Marian, 2002). The significance of this thesis resides in its capacity to support and eventually implement the development

of visually–guided reaching on the priori achieved foundation of motor programs for eye–movements (see discussion in Section 7.1.4).

3.2   Neurobiological bases of visuomotor coordination

In Section 3.1.2 we made the case that sensorimotor coupling is a fundamental operation that a neural system has to implement. While the unitary nature of the perception–action cycle is a ubiquitous feature of simple nervous systems, the sophistication of cortical circuitry in the primate's brain poses the problem of where and how perception and action are integrated. In this thesis, we address the issue of direct sensory control of action, defined as the immediate motor response evoked by the visual activity that codes the task parameters. We will try to argue that despite the sophistication of the circuitry, this process relies on a series of basic computational mechanisms. In this section, the most recent experimental data on the neurobiology of reaching movements is reviewed in search of the basic computational building–blocks of perception–action coupling.

Experimental results on arm control indicate that no cortical area is uniquely responsible for reaching (Andersen et al., 1997; Kalaska et al., 1997). The *distributed representation* of sensory and motor information and the co–existence of several levels of representation suggest that multiple cortical and subcortical sites simultaneously interact to produce an effective reaching to visual targets (Kalaska et al., 1992; MacKay, 1995; Caminiti et al., 1998). This cooperative interaction favors the idea of a more *parallel organization* of visuomotor transformation. This contrasts with the classical view of a sequential cascade of anatomically segregated neural populations, each generating a serial representation to be integrated in the next stage (see Figure 3.1, Kalaska et al., 1992).

Based on the fact that visually derived information is available in motor areas (Johnson et al., 1999) and signals coding the intention of movement have been observed in visual areas (Andersen et al., 1997), the visuomotor mapping can be realized by the simultaneous (as opposed to serial) engagement of neural populations distributed along the frontal and parietal lobes. These cell populations combine in a nonlinear way information coming from different input sources and are able to provide an output in a different coordinate system (Salinas and Abbott, 1995). In other words, visuo–to–motor mapping can be realized at multiple stages, in a widely distributed manner, and based on the activity of integrative mechanisms that function at the level of single cell or small–networks of neurons. In the

following, we shall see how these mechanisms are implemented in the brain for the control of arm–reaching movements.

3.2.1   Gradient architecture of parieto–frontal network

Visually guided reaching can be achieved through a combination of different sources of information, relating to target location, gaze direction, arm position and movement direction. Signals about the location of the target on the retina, and the position and movement of the eye and the arm appear to be distributed and co–exist in many reach–related areas from the parietal and frontal cortices (Johnson, 1992). The parietal cortex is known to contribute to the formation of multimodal spatial representations in different coordinate frames, while frontal activity is related to processes involved in the selection and execution of motor actions (Kalaska et al., 1997; Flash and Sejnowski, 2001).

Recent physiological recordings in the superior parietal and the frontal lobes revealed a gradient–like distribution of neural functional properties (Johnson et al., 1996; Marconi et al., 2001). These findings describe a gradual rostro–caudal change of response properties across the premotor and primary motor cortex. This change ranges within the frontal lobe from higher–order visuospatial processes (more rostrally) to lower–order actuator–specific signals (more caudally) (Johnson et al., 1996; Shen and Alexander, 1997). The visual–to–somatic gradient in the frontal lobe is represented in the left part of the brain in Figure 3.2.

Johnson and co–workers (1996) found a symmetrical gradient for the superior parietal cortex: arm–movement and posture related activity tended to occur more rostrally, whilst signal–related activity was found more caudally in the medial intermediate parietal area (MIP) (see representation of right part of the brain in Figure 3.2). Moreover, cell populations with similar response properties in the parietal and frontal areas are interconnected by cortico–cortical projections. These links are perhaps necessary for the match of retinal-, eye- and hand-related information during visually guided reaching (Caminiti et al., 1998; Marconi et al., 2001).

Based on these findings, Burnod and colleagues (1999) proposed that cortical control of reaching is distributed over a 'network of networks' in the parieto–frontal lobes. These networks form a visual–to–somatic gradient presented in Figure 3.2 and schematized as follows: (1) target location in retinal and eye coordinates is signaled by neurons at the caudal and intermediate levels of superior parietal cortex and in the rostral part of the frontal lobe
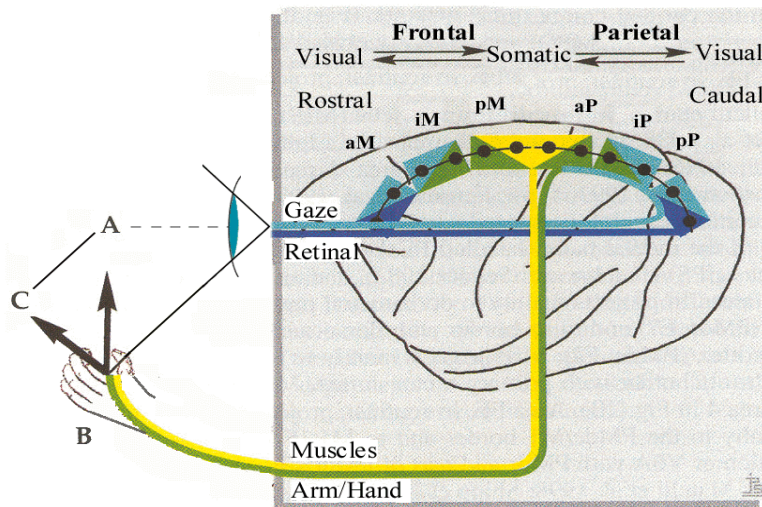
Figure 3.2: Parieto–frontal architecture with a visual–to–somatic gradient for visually guided reaching (adapted after Burnod et al., 1999). During reaching, the visuomotor transformation requires the combination of retinal (*dark blue*), gaze (*light blue*), arm/hand (*green*) and muscle (*yellow*) signals to move the hand from position B to the target in A (foveated) or in C (non-foveated). In right, cortical areas involved are represented. The authors grouped these areas approximately, in three parietal regions: anterior *aP*, intermediate *iP*, posterior *pP* and three frontal motor regions: anterior *aM*, intermediate *iM*, and posterior *pM* which are all reciprocally connected.

(see the dark blue lines and areas in figure); (2) eye–movement–related neurons (i.e., signal gaze direction) are distributed less caudally in the superior parietal cortex and less rostrally in the frontal lobe (indicated with light blue in figure); (3) neurons tuned to arm position extend more rostrally in the parietal lobe and more caudally in the frontal lobe (in green in figure); (4) arm movement-related activity predominate in the rostral bank of the central sulcus (in yellow in figure). At the border of these regions, where activities related to different signals overlap, combinatorial domains result with a role on integrating information from different sources (see Section 3.2.3).

3.2.2 Gain field mechanisms for coordinate transformation

The parieto–frontal network described above supports a gradual mapping of visual to somatic information through the involvement of neural populations coding different types of signals. In order to correctly guide motor action, the sensory signals coding retinal position

and gaze direction must be combined, as noted in Section 3.1.1 by an operation similar to a coordinate system transformation.

The most accepted perspective on of how this nonlinear mapping is implemented in the brain is that it is realized in a distributed manner, with multiple coordinate systems and reference frames forming the parietal cortex for spatial representation. Mapping of visual signals to motor output is realized by the smooth transition of information from one reference frame (or combinatorial domain) to another (Salinas and Abbott, 1995; Andersen et al., 1997; Caminiti et al., 1998; Snyder, 2000).

Andersen and collaborators (1997, 2000) have provided important insights on how neural circuits may implement nonlinear functions similar to coordinate transformation. They have suggested that the transformation from retina–centered coordinates to head– or body–centered coordinates can be understood in terms of *gain fields* associated with neurons in area 7a of the posterior parietal cortex. Cells in this area receive a convergence of eye position signals and visual signals and the response of a neuron is a product of the receptive field and the linear gain field. This multiplication implements a nonlinear function that can be described as a coordinate system transformation.

The gain field mechanism represents space in a distributed format, allowing inputs from multiple sensory systems with different spatial frames and outputting signals for action in several motor coordinate frames. Several neural network models based on gain field mechanisms have been proposed for the conversion of eye and retinal position into head–centered coordinates (Zipser and Andersen, 1988; Salinas and Abbott, 1995), or retina (eye–centered) and auditory (head–centered) signals into motor error coordinates (Andersen et al., 1997).

The solution to the nonlinear mapping problem illustrated above, focuses on the formation of distributed representations in multiple coordinate frames, by using gain field modulation. Another modality to combine signals for arm reaching, for instance the target location and the limb starting position is to converge them onto single cells. These then have to perform a match of the stimulus attributes (target location) with attributes of the appropriate motor response (joint angles to reach the target) (Kalaska et al., 1997). This approach is based on neural combinatorial properties and it is presented in the section below.

3.2.3   Combinatorial properties of neurons

The combinatorial properties of neurons have been observed in many areas of the parietal and frontal lobes (for a review see Wise et al., 1997; Caminiti et al., 1998). For instance, discharge rates of neurons in premotor areas and in the 7m parietal area co-vary with the direction of both eye and arm movement (Kalaska et al., 1997). Parietal cells in V6a (dorso–medial part of the parieto–occipital area) exhibit complex interactions between visual inputs and motor functions (Battaglia-Mayer et al., 2000). Moreover, about 40% of neurons in the premotor ventral area are modulated by the direction of gaze (Mushiake et al., 1997).

Based on these findings, Burnod and coworkers (1999) attempted to explain the biology underpinning sensorimotor transformation in terms of *combinatorial domains* and *matching units*. The authors proposed that the gradient architecture of the parieto–frontal network (see Section 3.2) favors the functional overlap of the regions coding for different reach–related signals (i.e., retinal, gaze, arm position, and muscle output). Signals from different sources along the visual–to–somatic axis can be matched within three combinatorial domains (see Figure 3.2):

- The anterior parietal *aP* and posterior frontal *pM* domain combines information on *muscle dynamics* with *positional and directional arm* information (c-domain1 combines yellow and green signals in Figure 3.2).

- The intermediate parietal *iP* and frontal motor area *iM* domain relates information on *movement direction* with *gaze positional and directional* signals (c-domain2 combines green and light blue signals),

- The posterior parietal *pP* and anterior frontal motor *pM* domain relates *positional and directional information* of gaze and arm with *visual inputs on the retina* (c-domain3 combines light and dark blue signals).

Within each domain, information can be combined along two other axes: the *position–direction* axis and the *sensory–motor* axis. It is well known that directional information is encoded in the activity of neurons in all reach–related areas from the parietal and frontal lobes (Georgopoulos et al., 1993; see Section 2.1.3). Hence, on the position–direction axis, processing units are tuned to both position and direction. That is, they respond maximally to movement in a certain direction and at a certain position in space.

Finally, neurons sharing the same combinatorial domain and similar position and directional tuning properties may have different temporal relationships to the signals relevant to reaching. This results in the existence of four types of units on the sensory–motor axis: *sensory units* that are timed–locked to sensory signals in all domains; *motor units*, time–locked to motor events; *matching units*, which learn correlation between sensory and motor signals; and *condition units*, which store correlation between sensorimotor signals and reinforcement contingencies.

The progressive match framework for visuomotor mapping

In our view, one of the best-articulated proposals of unsupervised visuomotor learning for arm-reaching was described in Burnod and co–workers (1999). This model exploits the gradient nature of the parieto–frontal architecture together with the combinatorial properties of neurons. In particular, it proposes that sensorimotor mapping is accomplished in a progressive manner, by the gradual involvement of sets of matching units belonging to different combinatorial domains. Synaptic learning results from the increase of the connection strength between two units, due to their repeated co–activation (e.g., Hebbian learning).

Below, we describe what in their view, are the first four stages required for the alignment of hand direction (in motor coordinates) toward the target position (in visual coordinates). Each stage corresponds to learning in different sets of matching units (for the description of the complete scenario see Burnod et al., 1999):

- *Motor babbling*. When the monkey moves the hand, co–activations in the first combinatorial domain (c-domain 1, see Section 3.2.3 above) due to inputs from motor commands and re–afferent somato-motor signals result in reinforced connections. After learning, these will allow the matching units to generate adequate muscles commands in order to produce an expected sensory effect.

- *Control of gaze and attention*. When the gaze shifts toward a stimulus, co–activations in the second combinatorial domain (c-domain 2) due to gaze movement and retinal input reinforce connections between neurons coding these signals. This learning enables the matching units to shift gaze to any focus of attention.

- *Hand tracking*. When the eyes look at the moving hand, co–activations in the third combinatorial domain (c-domain 3), due to inputs signaling gaze movement and hand

movement, result in reinforced connections that relate equivalent hand and gaze pathways. This set of matching units learns to perform visual tracking of hand movement.

- *Reaching to foveated targets.* When the hand moves toward the foveated target, co–activation of previously reinforced hand–gaze and gaze–retinal connections leads to learning in the connections of matching units. After training, these units will be able to serve reaching to foveated targets, even if the hand is not in the field of view.

Note that within the parieto–frontal network, gaze and arm directional and position signals play a unifying role. First, a subset of processing units is pre–selected based on gaze and arm position and these units then learn stable relationships between visually derived signals and somatomotor signals.

In summary, recent work on the coding of reaching movements in the brain has significantly advanced the knowledge of sensorimotor transformation on several fronts. Experimental evidence points to the fact that cortical control of reaching is distributed along a parieto–frontal network and it involves serial and parallel recruitment of neural populations distributed along a visual–to–somatic gradient. Coding of both arm and eye movements in an eye—centered reference frame, may represent the 'missing link' in understanding this visuomotor mapping.

Burnod and co–workers (1999) integrative framework was described in more detail, with the belief that it currently represents the most coherent proposal of how computational processes involved in reaching can be specified at the biological level. Up to the present, the simulation of a complete neural network to perform the neural operations of matching units has not been solved yet. In this thesis, we bring computational evidence from simulations with realistic spiking neural models for the learning mechanisms involved in the progressive match framework. In the remainder of this chapter, we focus upon discussing biologically inspired neural network models of visuomotor coordination.

3.3   Biologically inspired models of visuomotor mapping

At present, after several decades of applied research, the adaptive capabilities of motion control on biological organisms are still vastly superior to the capabilities of robotic systems. Evident sources of inspiration for the creation of more intelligent robots are the real nervous systems. In the recent years, numerous neural network models have been developed to

apply biologically inspired control mechanisms to various robot control tasks (Bullock et al., 1995; Zeller et al., 1995). Accordingly, several recent proposals have been made for a more physiologically inspired modeling of visuomotor coordination development (Burnod et al., 1992; Salinas and Abbott, 1995). Efforts on this direction have been aimed at the construction of neural architectures that are completely founded upon anatomical evidence. This means, to use neural elements that correspond as closely as possible to known neural cell types and to apply biologically plausible learning (Bullock et al., 1995).

This research direction reflects a change in emphasis away from the more traditional approaches that model visuomotor development based on control theory formalisms (Massone, 1995; Jordan, 1996; Jordan and Wolpert, 2000). That is, it explores alternate learning principles to the error-correction mechanisms (see Mohl, 1993). The attention is oriented towards investigating the efficacy of the neocortex organizational principles, when they are applied to arm-reaching tasks. These principles are of an unsupervised nature and are the result of correlation–based associations (e.g., Hebbian learning).

At first, it was suggested that nonlinear sensorimotor transformations could be modeled as sets of linearized representations, and so become a simple linear problem (Bullock et al., 1993). Furthermore, Baraduc and Guigon (2002) demonstrated using rigorous mathematical apparatus, that linear transformations can be learned by Hebbian associations, so long as the training examples satisfy a regularity condition. For cases where the examples do not satisfy the regularity condition (e.g., non–uniform distribution of preferred attributes of neurons in a population) learning the transformation is possible only with a stochastic error–correcting learning rule (see discussion in Section 3.3.2).

This section first reviews the main developmental paradigm for learning visuomotor coordination. Then we briefly discuss the relation between self–organizing and error–correcting mechanisms and finally, we focus upon our main interest that of physiologically inspired models driven by means of unsupervised learning.

3.3.1   The developmental paradigm

A large number of models that learn inverse kinematics through experience use a *circular reaction learning* protocol. The circular reaction concept was introduced by Piaget (1969) and describes the following behavioral loop: as a child performs random, spontaneously generated movements of his arm, his eyes follow the arm's motion, thereby enabling the learning

of a transformation from a visual representation of arm position to a motor representation of the same arm position. Inspired by this developmental process, researchers in robotics have implemented a similar adaptive process for learning inverse kinematics. This process is based on an autonomously controlled behavioral cycle with two components: production and perception, and is referred to as *motor babbling* (Kuperstein, 1988; Ritter et al., 1992).

During a motor babbling stage, the control system endogenously generates random movements, which displace the arm to different postures, bringing the end–effector into view. For each configuration of the arm, the vision system locates the end–effector and correlations between joint variables and eye variables are learned. After the correlations are stored in connection weights, the visual input passes through the weight network to generate a predicted arm activation vector. The difference between the predicted and externally generated activation vectors is used to modify the weights network and after many babbles the model will learn to accurately reach to the foveated objects in the workspace (Kuperstein, 1988; Gaudiano and Grossberg, 1991).

In the last decade, it has been pointed out that an algorithm based exclusively on random movement generation–and–observation for learning the inverse kinematics has a number of drawbacks (Jordan, 1996; see Section 3.3.2 below). Despite this limitation, the *action–perception* cycle remains the *de facto* behavioral framework for building developmental models of eye–hand coordination.

The *circular reaction learning* is not the only concept that computational researchers have borrowed from developmental psychology in an attempt to create artificial systems capable of adaptive, on–line control of goal directed reaching. Kuniyoshi and coworkers proposed a developmental framework based on the concepts of *assimilation* and *accommodation*, *imitation* and *entrainment dynamics* (Kuniyoshi and Berthouze, 1998; Kuniyoshi et al., 1999; see also Smith and Thelen, 1993). The implementation of these notions allows an artificial system to bootstrap itself towards a higher complexity through embodied interaction dynamics with the environment. Acquisition of eye–hand coordination and cross–modality mapping are attributed to the emergence of ordered structure from: interaction between many sensory-motor processes, embodiment, and basic pre–existing behavior patterns (i.e., reflexes) (Berthouze and Kuniyoshi, 1998; Kuniyoshi and Berthouze, 1998).

The importance of a pre–existing repertoire of motor primitives such as primitive reflexes, to support the development of coordinated movement has been also emphasized by Metta and co–workers (1999). The authors proposed a developmentally plausible solution for the

emergence of eye–hand coordination. Instead of computing the kinematics required by the transformation of information in 3D visual coordinates to hand coordinates, they implemented a direct mapping between the eye–head motor plant and the arm motor plant. In this approach, the eye–head movement is controlled using force fields. Consequently, the position of the target point in space can be coded using motor commands, namely the commands that control the position of the head and that of the eye with respect to the head. This allows the arm's force–fields to be obtained through a *motor–to–motor* mapping, from eye–head force fields. On the human infants, this mapping is assembled through experience and it is possible thanks to the built–in motor reflexes and the sensory–triggered motions that provide an effective means for linking vision and proprioception.

3.3.2 Self–organizing and error–based mechanisms for direction mapping

The direct inverse modeling approach outlined in the previous section has a number of limitations and cannot properly control nonlinear redundant objects (Jordan, 1996). More effective algorithms have been proposed, which solve the inverse kinematics problem in two phases. During an initial *learning by doing* period, the forward kinematics from joint angles to arm posture are learned. In a second stage, the inverse kinematics from desired trajectory to joint–angle map are learned, by using error–correction mechanisms (Mel, 1991; Jordan, 1996).

The error–correction algorithms represent the most common approach in modeling sensory guided reaching. A more biologically plausible alternative is to use unsupervised learning based on Hebbian associations. It was pointed out that correlation–based association mechanisms can assure the alignment of sensorimotor maps for the correct transfer of information for reaching (Baraduc and Guigon, 2002). We illustrate here the application of these learning rules to the direction–mapping problem.

Direction–mapping learning is an effective approach for translating spatial trajectory to end–effector directions, which has received increased attention in recent years amongst researchers working on modeling of sensorimotor coordination (Bullock et al., 1993; Fiala, 1995; see also Section 2.3.2). Perhaps the most illustrative example of learning direction mapping for visually guided movements is the self–organizing DIRECT model proposed by Bullock and colleagues (1993).

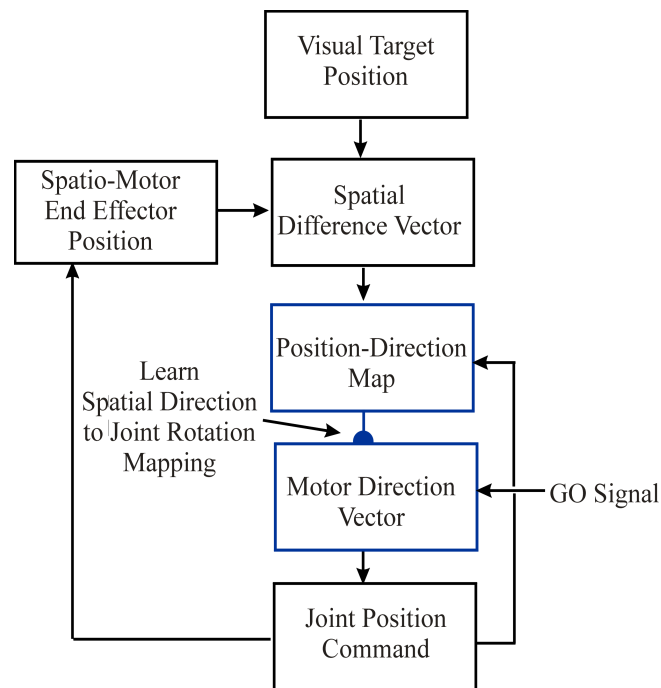The central operation in the DIRECT system is the mapping of spatial directions in body–

Figure 3.3: Processing stages of the self–organizing DIRECT model for control of movement (adapted after Bullock et al., 1993). Learning of the spatial direction to joint rotation mapping occurs during an action–perception cycle triggered by the GO signal, and is based on an improved Hebbian-like learning rule. See in text for details.

centered coordinates to joint directions in muscle–coordinates. This is accomplished in a number of processing stages as illustrated in Figure 3.3. First, the current position of the end effector is determined and the desired movement direction is computed as the difference between the visual representation of the target position and the actual representation of the end effector (first two stages in Figure 3.3). Second, the actual motor position signal is combined with the spatial direction vector using a self–organizing network. Learning at this stage results in the formation of a map where each cell is sensitive to a particular spatial direction in a particular position of joint space (the position–direction map in Figure 3.3).

Learning of spatial–to–motor direction mapping takes place in the third stage of the model (see the modifiable blue synapses in Figure 3.3). Here, translation of position–direction information into joint-directions is learned during a motor babbling period, when the GO signal generates random movements. The adaptation of the plastic synapses is based on a variant of normalized Hebbian learning mechanism, known as the outstar rule (Grossberg, 1998). After learning, the model is capable of visual control of reaching with tools and clamped joints, in the presence of visual input distortions or in the absence of visual signals

(i.e., blind reaching).

Learning of the differential transformation between the motor and spatial directions can also be implemented by using a gradient descent algorithm (Ritter et al., 1989). Fiala (1995) proposed an error–correction based algorithm that combines the DIRECT and VITE models (Bullock and Grossberg, 1989) in order to obtain bell–shaped velocity profiles of trajectories. In this architecture, the spatial directional vector is mapped to a motor direction vector through an intervening field of cells, referred to as direction mapping cells. The difference in activity between the spatial and motor direction vectors is computed and minimized with a gradient descent algorithm. Feedback information from the motor output layer is also taken into account during learning.

The model learns a kind of inverse Jacobian in a manner similar to that obtained by Ritter et al. (1989) and succeeds in reproducing straight–line motions and unimodal, bell–shaped velocity profiles. Based on the model's good performance, Fiala (1995) has tried to motivate the use of error–correction learning mechanisms with behavioral and biological evidence. Thus, behavioral results on spiraling–in movements during reaching suggest that the visual error between the hand and target is computed during movement (Roby-Brami and Burnod, 1995). However, these experimental findings also indicate that movement usually begins as a straight line along the transformed direction and a spiraling movement is often observed during the later part of the reaching. This observation suggests the existence of different stages in movement control, which, accordingly, may rely on different mechanisms.

According to Doya (1999), Hebbian synaptic mechanisms characterize the self–organization of the cerebral cortex, while error–correction learning is implemented in the cerebellum. Given the separate localization of these adaptation mechanisms a plausible hypothesis is that learning of movement control implies synaptic changes driven by different rules, in both structures. At the cortical level, correlation–based associations may be involved in the formation of stable connections for the alignment of visual and motor neural representations. This process allows the correct transfer of information to initiate the movement, by specifying the directional information. The involvement of error–correction mechanisms through the cerebellum may be necessary for the adjustment of the trajectory during the later stages of movement, to obtain an adequate reaching of the target (see also Fiala, 1995; Doya, 1999).

To conclude, there is currently a biologically motivated research direction in modeling of arm movement control. Much attention within this direction is given to the use of self–

organizing algorithms for learning of inverse kinematics. It has been argued that Hebbian association mechanisms can learn linear transformations and they can implement synaptic adaptation for development of visuomotor mapping. Furthermore, despite the limitations of direct inverse modeling, there is a growing research stream which builds–upon the developmental approach of visuomotor acquisition. In this thesis, modeling work of visuomotor mapping is based exclusively on unsupervised learning and it is placed within the developmental paradigm.

3.3.3   Models inspired by physiological data

A significant part of this chapter has been devoted to the review of recent experimental data on neurobiology of visuomotor transformation (Section 3.2). In comparison, the survey of biologically inspired modeling of sensorimotor development is less extensive. This is because, while much detail regarding the manner in which the brain solves the visuomotor transformation has been accumulated, the implementation of these concepts within artificial systems has only just begun. We believe that most importantly, the current progress towards the understanding of arm reaching neurobiology has given rise to a conceptual framework that explicates the neural basis of computation, allowing its readily implementation (see Section 3.2.3 for instance).

The advantages of constructing artificial systems inspired by real living systems are clear (Bullock et al., 1995; Cliff, 1995). Note that we are referring to an implementation fully grounded on physiological data, with architecture, neural models and learning rules all designed to be biologically plausible. We believe that an important step forward in this endeavor has been represented by the proposal of Burnod and colleagues (1999) (see Section 3.2.3). Even if the implementation of this model is not yet complete, it represents the result of a series of precursor models, which explore the alignment of hand, gaze and visual signal representations by unsupervised means. We review this previous work here.

Salinas and Abbott models

The biological modeling of the visuomotor transformation has been significantly advanced by the studies of Salinas and Abbott (1995, 1996). Their work addresses two major problems. The first problem considers how to find the mechanisms that ensure the alignment

of neural representations, in such a way that sensory activity representing target location evokes an appropriate motor response toward the target (the *alignment problem*). The second issue addresses the *coordinate transformation* problem. That is, if the target location in body–coordinates is determined by a combination of retinal position and gaze direction, then what set of weight connections assure that the correct combination is transferred to the motor network?

Their solution to the alignment problem consists of an unsupervised mechanism that transfers information between two networks composed of broadly selective neurons. Sensory neurons can encode retinal position solely, or retinal and gaze direction information, while motor neurons are assigned with preferred directions of movement. Both motor and sensory neurons firing rates are characterized by maximal cosine tuning curves (see Section 2.1.3 on direction coding).

The networks are coupled through modifiable synapses and learning occurs during the observation of the random movements generated by the motor network (e.g., motor babbling phase). The training procedure is based on a general Hebbian learning mechanism that modifies the synaptic weight w_{ij} between a j sensory neuron and an i motor neuron according with the correlation between their firing rates R_i^m and R_j^s:

$$w_{ij} =< R_j^s R_i^m > -k \tag{3.1}$$

where k is an arbitrary constant set to optimize the performance of the system. This process leads to an accurate alignment of the visual and motor representations, which ensures that sensory activity coding the target position is correctly used to guide the movement generated in the motor map. The authors derive a condition for the network alignment to occur. The condition is that the strength of the connection between a sensory neuron j and a motor neuron i depends on the magnitude of the difference between the neurons preferred locations.

Furthermore, the authors address the problem of linear (and nonlinear) mapping of the target location in retinal coordinates into movement direction in head–centered coordinates. It is shown that a sensory–motor network with synaptic weights that depends on the difference between the preferred direction of the motor neuron and the sum of the preferred retinal position and gaze direction for the sensory neuron in a pair can provide a general mechanism for linear coordinate transformation. The nonlinear coordinate transformation problem was further implemented with computational mechanisms at the population level,

which are shown to realize neural multiplication (Salinas and Abbott, 1996).

A more general computational framework, which deals in detail with the distributed representation of linear transformations and discusses the limitations of Hebbian mechanisms when compared to supervised means, in learning of nonlinear coordinate transformations can be found in Baraduc and Guigon (2002).

Burnod and colleagues work

A significant contribution to the understanding of computational mechanisms involved in the control of arm–reaching was brought by the work of Burnod and co–workers during the last decade. In comparison with Salinas and Abbott mathematical model, those of Burnod and colleagues are more tightly linked to the physiological and anatomical aspects of sensorimotor learning. Models developed by this research group incorporate somatic information beside the visually derived signals, in order to compute the appropriate motor command for reaching out to a visual target (Burnod et al., 1992; Baraduc et al., 1999).

A precursor to matching units' concept and neurons combinatorial properties was first described by Burnod et al. (1992). The primary goal of this study was to examine how the central nervous system might learn to reach towards a target position from any initial arm posture. The proposed architecture has three layers: (1) a map of *matching* units, which combine visual directional input and somatic arm signals and project to the next layer of synergy units; (2) a laterally connected map of *synergy* neurons which integrate the input from the matching layer and periphery feedback and which projects to a third (3) layer of *motor output* units that displace the hand in space. The appropriate combination of the visual information about movement trajectory with the kinesthetic information concerning the orientation of the arm in space is learned by the network from spontaneous movements.

Learning of the inverse kinematics from hand space to actuator space takes place in the matching units layer, according to correlation–based adaptation rules applied to sensory and motor co–activated events. The neural response in each layer is given by a combination of the inputs coming from the afferent source and the lateral signals from other units in the layer. The computation learned by the units can be approximated by a bilinear operation. This is interpreted by the authors as a projection of the visual information on a motor reference frame, that simultaneously rotates with the arm (Burnod et al., 1992). Once the network is trained, the signal on the direction of intended hand movement is projected onto

the matching units and activates the appropriate synergy units as a function of the starting arm posture.

A simpler, but more readily implementable variant of the above model was proposed recently by Baraduc et al. (1999). They consider a neural network model which combines broadly tuned (muscular) proprioceptive and visually derived directional input to calculate the angular motor commands for the initial part of the movement of a two–link arm. The arm position is represented by a set of 20 proprioceptive neurons coding muscle length. Direction of the desired movement is coded by a population of 50 visual neurons with cosine tuning functions. Motor commands are specified by a population of 50 motor neurons with optimal tuning property, which command motor synergies (direction in joint space). Compared to Burnod et al. (1992) model, this architecture is simplified by considering only two layers and learning only at one level. The advantage is that, in this case, the motor synergy layer encodes information directly in joint coordinates.

Information concerning the position of the arm and the visual desired direction is combined in two steps. First, an intermediate representation of arm position is formed in a somatic layer, by the combination of feed–forward proprioceptive signals and lateral intra-layer feedback. Second, activity in the somatic layer is combined with the visual directional activity. Only weights between the proprioceptive input layer and the somatic map are subject of learning. Training is performed by motor babbling in 5 positions and takes place according to a variant of the delta rule (i.e., supervised learning). After learning, the network generates arm commands (in joint coordinates) which move in the same direction as the visual input, over a large part of the visual inputs (16 directions of movements were tested).

Compared to previous models this approach has several advantages. In contrast to the DI-RECT model (Bullock et al., 1993), Baraduc and co–workers use broadly tuned motor neurons for learning of visuomotor mapping of direction. Compared with Salinas and Abbott approach, generation of movement is modeled with respect to the arm position (proprioceptive feedback). However, the somatic map organization produces a motor population vector, which largely deviates from the desired direction of movement. We believe this result might be improved by separating the motor map organization from learning the direction mapping task. A further discussion of these models can be found in the final chapter, along with an evaluation of the results of our work.