Multimedia Contents

77. Neurorobotics: From Vision to Action

Patrick van der Smagt, Michael A. Arbib, Giorgio Metta

The lay view of a robot is a mechanical human, and thus robotics has always been inspired by attempts to emulate biology. In this chapter, we extend this biological motivation from humans to animals more generally, but with a focus on the central nervous systems in its relationship to the bodies of these creatures. In particular, we investigate the sensorimotor loop in the execution of sophisticated behavior. Some of these sections concentrate on cases where vision provides key sensory data. Neuroethology is the study of the brain mechanisms underlying animal behavior, and Sect. 77.2 exemplifies the lessons it has to offer robotics by looking at optic flow in bees, visually guided behavior in frogs, and navigation in rats, turning then to the coordination of behaviors and the role of attention. Brains are composed of diverse subsystems, many of which are relevant to robotics, but we have chosen just two regions of the mammalian brain for detailed analysis. Section 77.3 presents the cerebellum. While we can plan and execute actions without a cerebellum, the actions are no longer graceful and become uncoordinated. We reveal how a cerebellum can provide a key ingredient in an adaptive control system, tuning parameters both within and between motor schemas. Section 77.4 turns to the mirror system, which provides shared representations which bridge between the execution of an action and the observation of that action when performed by others. We develop a neurobiological model of how learning may forge mirror neurons for hand movements, provide a Bayesian view of

a robot mirror system, and discuss what must be added to a mirror system to support robot imitation. We conclude by emphasizing that, while neuroscience can inspire novel robotic designs, it is also the case that robots can be used as embodied test beds for the analysis of brain models.

Part G | 77

77.1 Definitions and History

Neurorobotics may be defined as:

> *the design of computational structures for robots inspired by the study of the nervous systems of humans and other animals.*

We note the success of (deep) *artificial* neural networks – networks of simple computing elements whose connections change with *experience* – as providing a medium for parallel adaptive computation that has seen application in robot vision systems and controllers but here we emphasize neural networks derived from the study of specific *neurobiological* systems. Neurorobotics has a twofold aim: creating better machines which employ the principles of natural neural computation; and using the study of bio-inspired robots to improve understanding of the functioning of the brain. Chapter 75, *Biologically Inspired Robots*, complements our study of *brain design* with work on *body design*, the design of robotic control and actuator systems based on careful study of the relevant biology.

77.1.1 History and Definitions

Science has long been playing with technical replicas of biological behavior. As a famous example, *Walter* [77.1] described two *biologically inspired* robots, the electromechanical tortoises *Machina speculatrix* and *M. docilis* (though each body has wheels, not legs). *M. speculatrix* has a steerable photoelectric cell, which makes it sensitive to light, and an electrical contact, which allows it to respond when it bumps into obstacles. The photoreceptor rotates until a light of moderate intensity is registered, at which time the organism orients itself towards the light and approaches it. However, very bright lights, material obstacles, and steep gradients are repellent to the *tortoise*. The latter stimuli convert the photoamplifier into an oscillator, which causes alternating movements of butting and withdrawal, so that the robot pushes small objects out of its way, goes around heavy ones, and avoids slopes. The *tortoise* has a *hutch*, which contains a bright light. When the machine's batteries are charged, this bright light is repellent. When the batteries are low, the light becomes attractive to the machine and the light continues to exert an attraction until the tortoise enters the hutch, where the machine's circuitry is temporarily turned off until the batteries are recharged, at which time the bright hutch light again exerts a negative tropism. The second robot, *M. docilis* was produced by grafting onto *M. speculatrix* a circuit designed to form conditioned reflexes. In one experiment, Walter connected this circuit to the obstacle-avoiding device in *M. speculatrix*. Training consisted of blowing a whistle just before bumping the shell.

Although Walter's controllers are simple and not based on neural analysis, they do illustrate an attempt to gain inspiration from seeking the simplest mechanisms that will yield an interesting class of biologically inspired robot behaviors, and then showing how different additional mechanisms yield a variety of enriched behaviors. *Braitenberg*'s book [77.2] is very much in this spirit and has entered the canon of neurorobotics. While their work provides a historical background for the studies surveyed here, we instead emphasize studies inspired by the computational neuroscience of the mechanisms serving vision and action in the human and in animal brains. We seek lessons from linking behavior to the analysis of the internal workings of the brain (1) at the relatively high level of characterizing the functional roles of specific brain regions (or the functional units of analysis called schemas, Sect. 77.2.4), and the behaviors which emerge from the interactions between them, and (2) at the more detailed level of models of neural circuitry linked to the data of neuroanatomy and neurophysiology. There are lessons for neurorobotics to be learned from even finer-scale analysis of the biophysics of individual neurons and the neurochemistry of synaptic plasticity, but these are beyond the scope of this chapter (see *Segev* and *London* [77.3] and *Fregnac* [77.4], respectively, for entry points into the relevant computational neuroscience).

The plan of this chapter is as follows. We will start with explaining how the higher-level *cognitive* functionality of vision-based planning and navigation is realized in biology, and how this relates to robotic systems (Sect. 77.2). We then (Sect. 77.3) explain vertebrate movement generation itself, and put forth a theory on what role the cerebellum plays in tuning and coordinating actions. This is followed by a section on the mirror system and its roles in action recognition and imitation (Sect. 77.4). The extroduction will then invite readers to explore the many other areas in which neurorobotics offers lessons from neuroscience to the development of novel robot designs. What follows, then, can be seen as a contribution to the continuing dialog between robot behavior and animal and human behavior in which particular emphasis is placed on the search for the neural underpinnings of vision, visually guided action, and cerebellar control.

77.2 The Case for Vision

Before we turn to vertebrate brains for much of our inspiration for neurorobotics, we briefly sample the rich literature on insect-inspired research. Among the founding studies in computational neuroethology were a series of reports from the laboratory of Werner Reichardt in Tübingen, which linked the delicate anatomy of the fly's brain to the extraction of visual data needed for flight control. More than 40 years ago, *Reichardt* [77.5] published a model of motion detection inspired by this work that has long been central to discussions of visual motion in both the neuroscience and robotics literatures. *Borst* and *Dickinson* [77.6] provide a recent study of continuing biological research on visual course control in flies. Such work has inspired a large number of robot studies, including those of *van der Smagt* and *Groen* [77.7], *van der Smagt* [77.8], *Liu* and *Usseglio-Viretta* [77.9], *Ruffier* et al. [77.10], and *Reiser* and *Dickinson* [77.11].

77.2.1 Optic Flow in Bees and Robots

Here, however, we look in a little more detail at honeybees. *Srinivasan* et al. [77.15] continued the tradition of studying image motion cues in insects by investigating how optic flow (the flow of pattern across the eye induced by motion relative to the environment) is exploited by honeybees to guide locomotion and navigation. They analyzed how bees perform a smooth
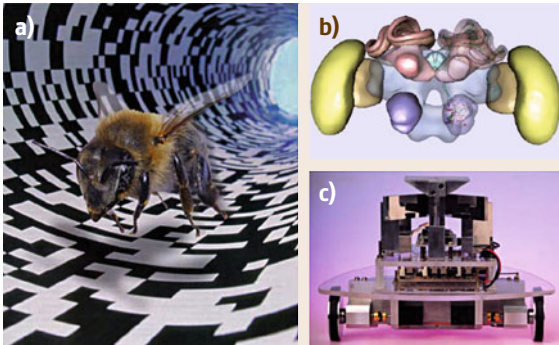

Fig. 77.1 (a) Observation of the trajectories of honeybees flying in visually textured tunnels has provided insights into how bees use optic flow cues to regulate flight speed and estimate distance flown, and balance optic flow in the two eyes to fly safely through narrow gaps (images courtesy of *Srinivasan* et al. [77.12]). This information has been used to build autonomously navigating robots. **(b)** Schematic illustration of a honeybee brain, carrying about a million neurons within $\approx 1\,\text{mm}^3$ (after [77.13]). **(c)** A mobile robot guided by an optic flow algorithm based on the studies exemplified in [77.14]

landing on a flat surface: image velocity is held constant as the surface is approached, thus automatically ensuring that flight speed is close to zero at touchdown. This obviates any need for explicit knowledge of flight speed or height above the *ground*. This landing strategy was then implemented in a robotic gantry to test its applicability to autonomous airborne vehicles. *Barron* and *Srinivasan* [77.14] investigated the extent to which ground speed is affected by headwinds. Honeybees were trained to enter a tunnel to forage at a sucrose feeder placed at its far end (Fig.77.1a). The bees used visual cues to maintain their ground speed by adjusting their airspeed to maintain a constant rate of optic flow, even against headwinds which were, at their strongest, 50% of a bee's maximum recorded forward velocity.

Vladusich et al. [77.16] studied the effect of adding goal-defining landmarks. Bees were trained to forage in an optic-flow-rich tunnel with a landmark positioned directly above the feeder. They searched much more accurately when both odometric and landmark cues were available than when only odometry was available. When the two cue sources were set in conflict, by shifting the position of the landmark in the tunnel during tests, bees overwhelmingly used landmark cues rather than odometry. This, together with other such experiments, suggests that bees can make use of odometric and landmark cues in a more flexible and dynamic way than previously envisaged. In earlier studies of bees flying down a tunnel, *Srinivasan* and *Zhang* [77.17] placed different patterns on the left and right walls. They found that bees balance the image velocities in the left and right visual fields. This strategy ensures that bees fly down the middle of the tunnel, without bumping into the side walls, enabling them to negotiate narrow passages or to fly between obstacles. This strategy has been applied to a corridor-following robot (Fig. 77.1c). By holding constant the average image velocity as seen by the two eyes during flight, the bee avoids potential collisions, slowing down when it flies through a narrow passage. The movement-sensitive mechanisms underlying these various behaviors differ qualitatively as well as quantitatively, from those that mediate the optomotor response (e.g., turning to track a pattern of moving stripes) that had been the initial target of investigation of the Reichardt laboratory. The lesson for robot control is that flight appears to be coordinated by a number of visuomotor systems acting in concert, and the same lesson can apply to a whole range of tasks that must convert vision to action. Of *course*, vision is but one of the sensory systems that play a vital role in insect behavior. *Webb* [77.18] uses her own work on robot

design inspired by the auditory control of behavior in crickets to anchor a far-ranging assessment of the extent to which robotics can offer good models of animal behaviors.

77.2.2 Visually Guided Behavior in Frogs and Robots

Lettvin et al. [77.19] treated the frog's visual system from an ethological perspective, analyzing circuitry in relation to the animal's ecological niche to show that different cells in the retina and the visual midbrain region known as the tectum were specialized for detecting predators and prey. However, in much visually guided behavior, the animal does not respond to a single stimulus, but rather to some property of the overall configuration. We thus turn to the question *what does the frog's eye tell the frog?*, stressing the embodied nervous system or, perhaps equivalently, an action-oriented view of perception. Consider, for example, the snapping behavior of frogs confronted with one or more fly-like stimuli. *Ingle* [77.20] found that it is only in a restricted region around the head of a frog that the presence of a fly-like stimulus elicits a snap, that is, the frog turns so that its midline is pointed at the stimulus and then lunges forward and captures the prey with its tongue. There is a larger zone in which the frog merely orients towards the target, and beyond that zone the stimulus elicits no response at all. When confronted with two *flies* within the snapping zone, either of which is vigorous enough that it could elicit a snapping response alone, the frog exhibits one of three reactions: it snaps at one of the flies, it does not snap at all, or it snaps in between at the *average fly*. *Didday* [77.21] offered a simple model of this choice behavior which may be considered as the prototype for a *winner-take-all* (WTA) model, which receives a variety of inputs and (under ideal circumstances) suppresses the representation of all but one of them; the one that remains is the *winner* that will play the decisive role in further processing. This was the beginning of *Rana computatrix* (see *Arbib* [77.22, 23] for overviews).

Studies on frog brains and behavior inspired the successful use of potential fields for robot navigation strategies. Data on the strategies used by frogs to capture prey while avoiding static obstacles (*Collett* [77.24]) grounded the model by *Arbib* and *House* [77.25], which linked systems for depth perception to the creation of spatial maps of both prey and barriers. In one version of their model, they represented the map of prey by a potential field with long-range attraction and the map of barriers by a potential field with short-range repulsion, and showed that summation of these fields yielded a field that could guide the frog's de-

tour around the barrier to catch its prey. *Corbacho* and *Arbib* [77.26] later explored a possible role for learning in this behavior. Their model incorporated learning in the weights between the various potential fields to enable adaptation over trials as observed in the real animals. The success of the models indicated that frogs use reactive strategies to avoid obstacles while moving to a goal, rather than employing a planning or cognitive system. Other work, *Cobas* and *Arbib* [77.27], studied how the frog's ability to catch prey and avoid obstacles was integrated with its ability to escape from predators. These models stressed the interaction of the tectum with a variety of other brain regions such as the pretectum (for detecting predators) and the tegmentum (for implementing motor commands for approach or avoidance).

Arkin [77.28] showed how to combine a computer vision system with a frog-inspired potential field controller to create a control system for a mobile robot that *could* successfully navigate in a fairly structured environment using camera input. The resultant system thus enriched other roughly contemporaneous applications of potential fields in path planning with obstacle avoidance for both manipulators and mobile robots (*Khatib* [77.29], *Krogh* and *Thorpe* [77.30]). The work on *Rana computatrix* proceeded at two levels – both biologically realistic neural networks and in terms of functional units called *schemas*, which compete and cooperate to determine behavior. Section 77.2.4 will show how more general behaviors can emerge from the competition and cooperation of perceptual and motor schemas, as well as more abstract coordinating schemas. Such ideas were, of course, developed independently by a number of authors, and so entered the robotics literature by various routes, of which the best known may be the subsumption architecture of *Brooks* [77.31] and the ideas of Braitenberg cited above, whereas *Arkin's* work on behavior-based robotics [77.32] is, indeed, rooted in schema theory. *Arkin* et al. [77.33] present a recent example of the continuing interaction between robotics and ethology, offering a novel method for creating high-fidelity models of animal behavior for use in robotic systems based on a behavioral systems approach (i. e., based on a schema-level model of animal behavior, rather than analysis of biological circuits in animal brains), and describe how an ethological model of a domestic dog can be implemented with AIBO, the Sony entertainment robot.

77.2.3 Navigation in Rat and Robot

The tectum, the midbrain visual system which determines how the frog turns its whole body towards it prey or orients it for escape from predators (Sect. 77.2.2), is

homologous with the superior colliculus of the mammalian midbrain. The rat superior colliculus has been shown to be *frog like*, mediating approach and avoidance (*Dean* et al. [77.34]), whereas the best-studied role of the superior colliculus of cat, monkey, and human is in the control of saccades, rapid eye movements to acquire a visual target. Moreover, the superior colliculus can integrate auditory and somatosensory information into its visual frame (*Stein* and *Meredith* [77.35]), and this inspired *Strosslin* et al. [77.36] to use a biologically inspired approach based on the properties of neurons in the superior colliculus to learn the relation between visual and tactile information in control of a mobile robot platform. More generally, then, the comparative study of mammalian brains has yielded a rich variety of computational models of importance in neurorobotics. In this section, we further introduce the study of *mammalian neurorobotics* by looking at studies of mechanisms of the rat brain for spatial navigation.

The frog's detour behavior is an example of what *O'Keefe* and *Nadel* [77.37] called the *taxon (behavioral orientation) system* (as in *Braitenberg*, [77.38] a *taxis* (plural *taxes*) is an organism's response to a stimulus by movement in a particular direction). They distinguished this from a system for *map-based navigation* and proposed that the latter resides in the hippocampus, though *Guazzelli* et al. [77.39] qualified this assertion, showing how the hippocampus may function as *part of* a cognitive map. The taxon versus map distinction is akin to the distinction between reactive and deliberative control in robotics (*Arkin* et al. [77.33]). It will be useful to relate taxis to the notion of an *affordance* (*Gibson* [77.40]), a feature of an object or environment relevant to action, for example, in picking up an apple or a ball, the identity of the object may be irrelevant, but the size of the object is crucial. Similarly, if we wish to push a toy car, recognizing the make of car copied in the toy is irrelevant, whereas it is crucial to recognize the placement of the wheels to extract the direction in which the car can be readily pushed. Just as a rat may have basic taxes for approaching food or avoiding a bright light, say, so does it have a wider repertoire of affordances for possible actions associated with the immediate sensing of its environment. Such affordances include *go straight ahead* for visual sighting of a corridor, *hide* for a dark hole, *eat* for food as sensed generically, *drink* similarly, and the various turns afforded by, e.g., the sight of the end of the corridor. It also makes rich use of olfactory cues. In the same way, a robot's behavior will rely on a host of reactions to local conditions in fulfilling a plan, e.g., knowing that it must go to the end of a corridor it will nonetheless use local visual cues to avoid hitting obstacles or to determine through which angle to turn when reaching a bend in the corridor.

Both normal and hippocampal-lesioned rats can learn to solve a simple T-maze (e.g., learning whether to turn left or right to find food) in the absence of any consistent environmental cues other than the T-shape of the maze. If anything, the lesioned animals learn this problem faster than normal ones. After the criterion was reached, probe trials with an eight-arm radial maze were interspersed with the usual T-trials. Animals from both groups consistently chose the side to which they were trained on the T-maze. However, many did not choose the 90° arm but preferred either the 45° or 135° arm, suggesting that the rats eventually solved the T-maze by learning to rotate within an egocentric orientation system at the choice point through *approximately* 90°. This leads to the hypothesis of an *orientation vector* being stored in the animal's brain but does not tell us where or how the orientation vector is stored. One possible model would employ coarse coding in a linear array of cells, coding for turns from $-180°$ to $+180°$. From the behavior, one might expect that only the cells close to the preferred *behavioral* direction are excited, and that learning *marches* this peak from the old to the new preferred direction. To *unlearn* $-90°$, say, the array must reduce the peak there, while at the same time *building* a new peak at the new direction of $+90°$. If the old peak has *massp(t)* and the new peak has *massq(t)*, then as $p(t)$ declines toward 0 while $q(t)$ increases steadily from 0, the center of mass will progress from $-90°$ to $+90°$, fitting the behavioral data.

The determination of movement direction was modeled by *rat-ification* of the *Arbib* and *House* [77.25] model of frog detour behavior. There, prey was represented by excitation coarsely coded across a population, while barriers were encoded by inhibition whose extent closely matched the retinotopic extent of each barrier. The sum of excitation was passed through a winner-takes-all circuit to yield the choice of movement direction. As a result, the direction of the gap closest to the prey, rather than the direction of the prey itself, was often chosen for the frog's initial movement. The same model serves for behavioral orientation once we replace the direction of the prey (frog) by the direction of the orientation vector (rat), while the barriers correspond to the presence of walls rather than alley ways.

To approach the issue of how a cognitive map can extend the capability of the affordance system, *Guazzelli* et al. [77.39] extended the *Lieblich* and *Arbib* [77.41] approach to building a cognitive map as a *world graph*, a set of nodes connected by a set of edges, where the nodes represent recognized places or situations, and the links represent ways of moving from one situation to another. A crucial notion is that a place encountered in different circumstances may be repre-
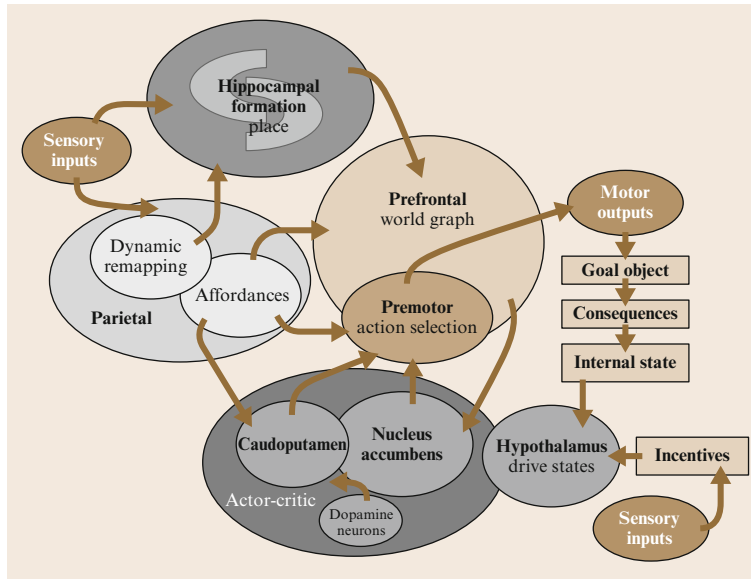
Fig. 77.2 The TAM-WG model has at its basis a system, TAM (the taxon affordance model), for exploiting affordances. This is elaborated by a system, WG (the world graph), which can use a cognitive map to plan paths to targets which are not currently visible. Note that the model processes two different kinds of sensory inputs. At the *bottom right* are those associated with, e.g., hypothalamic systems for feeding and drinking and that may provide both incentives and rewards for the animal's behavior, contributing both to behavioral choices, and to the reinforcement of certain patterns of behavior. The nucleus accumbens and caudoputamen mediate an actor-critic style of reinforcement learning based on the hypothalamic drive of the dopamine system. The sensory inputs at the *top left* are those that allow the animal to sense its relation with the external world, determining both where it is (the hippocampal place system), as well as the affordances for action (the parietal recognition of affordances can shape the premotor selection of an action). The TAM model focuses on the parietal–premotor reaction to immediate affordances; the WG model places action selection within the wider context of a cognitive map (after *Guazzelli* et al. [77.39])

sented by multiple nodes, but that these nodes may be merged when the similarity between these circumstances is recognized. They model the process whereby the animal decides where to move next, on the basis of its current drive state (hunger, thirst, fear, etc.). The emphasis is on spatial maps for guiding locomotion into regions not necessarily currently visible, rather than retinotopic representations of immediately visible space, and yields exploration and latent learning without the introduction of an explicit exploratory drive. The model shows:

1. How a route, possibly of many steps, may be chosen that leads to the desired goal.
2. How short cuts may be chosen.
3. Through its account of node merging why, in open fields, place cell firing does not seem to depend on direction.

The overall structure and general mode of operation of the complete model is shown in Fig. 77.2, which gives a vivid sense of the lessons to be learned by studying not only specific systems of the mammalian brain

but also their patterns of large-scale interaction. This model is but one of many inspired by the data on the role of the hippocampus and other regions in rat navigation. Here, we just mention as pointers the wider literature the papers by *Girard* et al. [77.42] and *Meyer* et al. [77.43], which are part of the *Psikharpax* project, which does for rats what *Rana computatrix* did for frogs and toads.

77.2.4 Salience and Visual Attention

Discussions of how an animal (or robot) grasps an object assume that the animal or robot is attending to the relevant object. Thus, whatever the subtlety of processing in the canonical and mirror systems for grasping, its success rests on the availability of a visual system coupled to an oculomotor control system that bring foveal vision to bear on objects to set the parameters needed for successful interaction. Indeed, the general point is that attention greatly reduces the processing load for animal and robot. The catch, of course, is that reducing the computing load is a Pyrrhic victory unless

the moving focus of attention captures those aspects of behavior that are relevant for the current task – or supports necessary *priority interrupts*. Indeed, directing attention appropriately is a topic for which there is a great richness of both neurophysiological data and robotic application (see *Deco* and *Rolls* [77.44] and *Choi* et al. [77.45]).

In their neuromorphic model of the bottom-up guidance of attention in primates, *Itti* and *Koch* [77.46] decompose the input video stream into eight feature channels at six spatial scales. After surround suppression, only a sparse number of locations remain active in each map, and all maps are combined into a unique *saliency map*. This map is scanned by the focus of attention in order of decreasing saliency through the interaction between a winner-takes-all mechanism (which selects the most salient location) and an *inhibition-of-return* mechanism (which transiently suppresses recently attended locations from the saliency map). Because it includes a detailed low-level vision front-end, the model has been applied not only to laboratory stimuli, but also to a wide variety of natural scenes, predicting a wealth of data from psychophysical experiments.

When specific objects are searched for, low-level visual processing can be biased both by the gist (e.g., *outdoor suburban scene*) and also for the features of that object. This top-down modulation of bottom-up processing results in an ability to guide search towards targets of interest (*Wolfe* [77.47]). Task affects eye movements (*Yarbus* [77.48]), as do training and general expertise. *Navalpakkam* and *Itti* [77.49] propose a computational model which emphasizes four aspects that are important in biological vision: determining the task relevance of an entity, biasing attention for the low-level visual features of desired targets, recognizing these targets using the same low-level features, and incrementally building a visual map of task relevance

at every scene location. It attends to the most salient location in the scene, and attempts to recognize the attended object through hierarchical matching against object representations stored in long-term memory. It updates its working memory with the task relevance of the recognized entity and updates a topographic task-relevance map with the location and relevance of the recognized entity; for example, in one task the model forms a map of likely locations of cars from a video clip filmed while driving on a highway. Such work illustrates the continuing interaction between models based on visual neurophysiology and human psychophysics with the tackling of practical robotic applications.

Orabona et al. [77.50] implemented an extension of the Itti–Koch model on a humanoid robot with moving eyes, using log-polar vision as in *Sandini* and *Tagliasco* [77.51], and changing the feature construction pyramid by considering proto-object elements (blob-like structures rather than edges). The inhibition-of-return mechanism has to take into account a moving frame of reference, the resolution of the fovea is very different from that at the periphery of the visual field, and head and body movements need to be stabilized. The control of movement might thus have a relationship with the structure and development of the attention system. *Rizzolatti* et al. [77.52] proposed a role for the feedback projections from premotor cortex to the parietal lobe, assuming that they form a tuning signal that dynamically changes visual perception. In practice, this can be seen as an implicit attention system that *selects* sensory information while the action is being prepared and subsequently executed (*Flanagan* and *Johansson* [77.53], *Flanagan* et al. [77.54], and *Mataric* and *Pomplun* [77.55]). The early responses, before action onset, of many premotor and parietal neurons suggest a premotor mechanism of attention that deserves exploration in further work in neurorobotics.

77.3 Vertebrate Motor Control

The body of literature on primate motor control is, of course, vast, and gives a patchy view on the principles behind it. Getting a clear view of how limb and general body control functions is difficult; moreover, there are no clear proofs of whether any of the existing views on motor control are correct.

But there exist a few observations of the human central and peripheral nervous systems from which clear conclusions can be drawn. The first observation is the presence of neural communication delays. How does the system know the position of limbs? There are two principled methods: (1) through proprioceptive signals, consisting of muscle spindles and Golgi tendon or-

gans (GOs); and (2) through skin information. It is, however, not very likely that information from muscle spindles and GOs are accurate enough to code limb position. Tendon organs are sensitive to forces along in-series motor units and there is no physiological evidence that Golgi tendon organs signal muscle length (but, of course, force changes with muscle length, so during movement a correlation is found). There is another problem with respect to limb position, which is particularly clear for fingers: flexibilities and nonlinear relationships between finger position and muscle force, in combination with the imprecise receptors, makes the relationship between GO/spindle data and finger posi-

Table 77.1 Classification of sensory fibers from muscle (after [77.56])

Type	Receptor	Axon diameter [μm]	Transmission speed [m/s]	Sensitive to
Ia	Primary spindle endings	12—20	60—100	Muscle length and rate of change
Ib	Golgi tendon organs	12—20	60—100	Muscle tension
II	Secondary spindle endings	6—12	30—70	Muscle length
III	Free nerve endings	2—6	10—30	Skin signals

tion too complex and variable to be a likely candidate to code finger position; after all, the sensors are in the forearm rather than in the fingers; and information on finger position is not available in muscle movement or tendon force. Furthermore, muscle spindle data is noisy [77.57]. It has conversely been shown [77.58, 59] that the receptors in hairy skin code information that can be related to finger position; furthermore, similar data have been found for the knee joint [77.60]. Table 77.1 lists nerve transmission speeds for these signals.

Since neurons are only to be found in the spinal cord and the brain, for hand skin, therefore, we can expect signal transfer delays to the spine of around 30—100 ms. Round-trip muscle activation is, therefore, around 70 ms for signals based on skin data [77.61], or around 25 ms for spindle-based signals (we have verified these delays by measuring hand skin-based reflexes by measuring the corresponding electromyography(EMG) signals, and found a round-trip delay of around 75 ms. Spindle-based feedback for the wrist was measured at around 25 ms.

Of course, when a sensory signal has to be processed in the brain, the delays are correspondingly longer. At any rate: error-correcting feedback control has delays of several tens of milliseconds; feedback control based on such delays cannot lead to any acceptable accuracy with the movement speeds that humans typically display. This means that large portions of our movement, over time frames in the order of 100 ms or more, need to be controlled open loop.

A second important observation is our generalizing capabilities. Consider the case of playing fast and accurate sports, e.g., table tennis. During play, we obtain sensory visual, haptic, and tactile sensory data, the result of which must lead to an accurate movement of the bat in order to score a point. Even a player with little training is able to do this rather accurately: at ball flight times between 200 and 500 ms, the brain does not have much time to plan an accurate whole-body movement for each and every possible sensory state, but we are usually capable of returning the ball. Training helps, but we do not need to exhaustively learn many states in the very high-dimensional sensor space in which our observations move.

Generalization can only be done with reasonably accurate models of the sensor/motor behavior. How-ever, models of our motor system are difficult to obtain: variations such as including payloads, wearing heavy clothing, muscle fatigue, etc., do not influence our accuracy considerably.

77.3.1 The Flat Hierarchy of Neurocontrol

How is an open-loop movement generated? In this paper we concentrate on voluntary vertebrate motor control; the only reason for any animal to have a brain is to generate movement. Moreover, despite differences in brain structures, there is a large correspondence in movement patterns among the whole animal kingdom, irrespective of the presence of a cortical structure or a cerebellum. What parts of the brain are directly involved in movement?

The major role of the cerebral cortex seems to be unsupervised learning to establish relationships between sensory and action patterns [77.62]. The neocortex is only to be found in mammals; experiments with decorticated cats [77.63] clearly show that the cortex is not necessary to generate movement; rather, it is likely that the motor cortex models and weighs movements, to subsequently make decisions based thereon.

The major role of the basal ganglia seems to be reinforcement learning to filter out unwanted movements [77.62]. They play a dominant role in movement generation or gating (filtering) of generated movement patterns. The effect of Parkinson's disease (the inability to initiate movement) and Huntington's disease (the inability to prevent unwanted movement) on the basal ganglia is well known and clearly indicates their function.

The major role of the cerebellum seems to be supervised learning of motor patterns [77.62]. Moreover, decerebellation does not lead to complete movement loss. An individual with cerebellar lesions may be able to move the arm to successfully reach a target and to successfully adjust the hand to the size of an object. However, the action cannot be made swiftly and accurately, and the ability to coordinate the timing of the two subactions is lacking. The behavior will thus exhibit decomposition of movement – first the hand is moved until the thumb touches the object, and only then is the hand shaped appropriately to grasp the object [77.64].

Robot control usually favors a strict hierarchical approach. A typical robot works as follows. At the low-

est level, a very fast ($\approx 100\,\mu$s) current control loop controls the rotation of the dc motor. On top of that, a torque controller (running typically at 1 kHz) controls the torque of all joints, and is in its turn controlled by an impedance or position controller. On top of that, typically, a Cartesian path planner forms the slowest loop. An error in any of these elements will disable the robot.

A look into evolutionary development of neural control thus makes immediately clear that a strict hierarchical approach is not viable in neural control. Although any of the above-mentioned brain regions is important in movement control, and similar structures can be found in any vertebrate, their dysfunction leads to movement degradation but not to movement loss (this is, of course, not true for the spinal cord, which (combines and) transmits the controls to the muscles). Also, the development of the neural system shows that animals were always capable of movement – irrespective of their brain structure. However, the cerebellum is usually rightly focused upon when analyzing vertebrate movement. How do the parts of the brain collaborate towards smooth goal-directed movement?

In placing the function of the cerebellum in the loop, a normal distinction is to consider the cerebellum as representing (a) a forward or direct model which represents the path from motor command to motor output, or (b) an inverse model of motor function, i. e., going from a desired motor outcome to a set of motor commands likely to achieve it. As we have just suggested, the action plan unfolds as if it were feedforward or open loop when the actual parameters of the situation match the stored parameters, while a feedback component is employed to counteract disturbances (current feedback) and to learn from mistakes (learning from feedback). This is obtained by relying on a forward model that predicts the outcome of the action as it unfolds in real time. The accuracy of the forward model can be evaluated by comparing the output generated by the system with the signals derived from sensory feedback (*Miall* et al. [77.65]). Also, delays must be accounted for to address the different propagation times of the neural pathways carrying the predicted and actual outcome of the action. Note that the forward model in this case is relatively simple, predicting only the motor output in advance; since motor commands are generated internally it is easy to imagine a predictor for these signals (known as an efference copy). The inverse model, on the other hand, is much more complicated since it maps sensory feedback (e.g., vision) back into motor terms.

We suggest a much simpler approach to the vertebrate control system. However, let us first look into the functionality of the lower-level apparatus: muscle, spinal cord, and cerebellum.

77.3.2 On Spinal Cord and Muscle

The key element in movement generation is given by two building blocks: (a) our muscles, and (b) the spinal cord. Muscle behavior is strongly nonlinear; the exerted force decreases nonlinearly with velocity (Fig. 77.3) and varies nonlinearly with length (Fig. 77.4).

Limb movement, however, is caused by a complex of muscles – for instance, the human arm uses a total of 19 muscle groups for planar motion of the elbow and shoulder alone (*Nijhof* and *Kouwenhoven* [77.67]) with altogether highly nonlinear dynamics. How can this large number of actuators be controlled without feedback error control?

The concept is simple and was first well described by *Bernstein* [77.68]: skeletal muscles are always controlled in functional groups, leading to *synergies* of movement. Rather than activating muscles indepen-
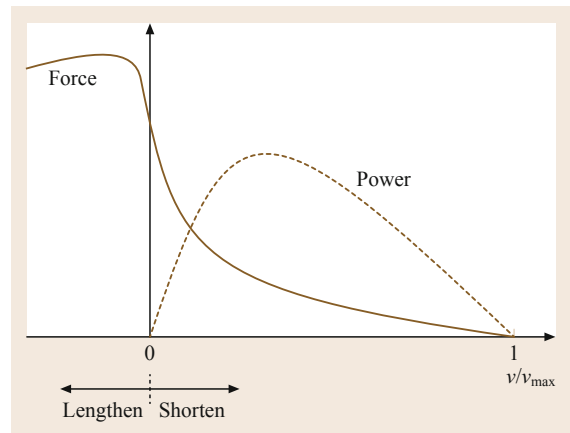

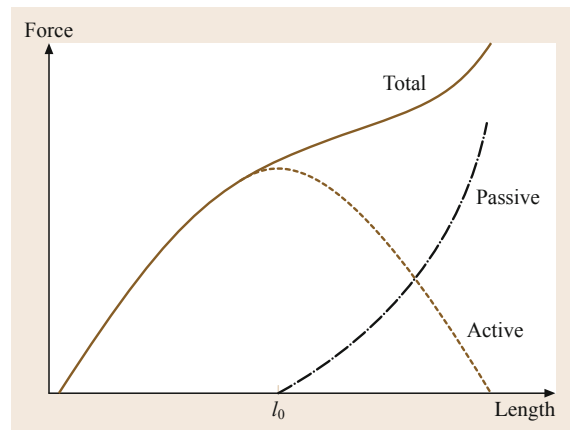**Fig. 77.3** The force/velocity and power/velocity relationship of muscle (after [77.56])

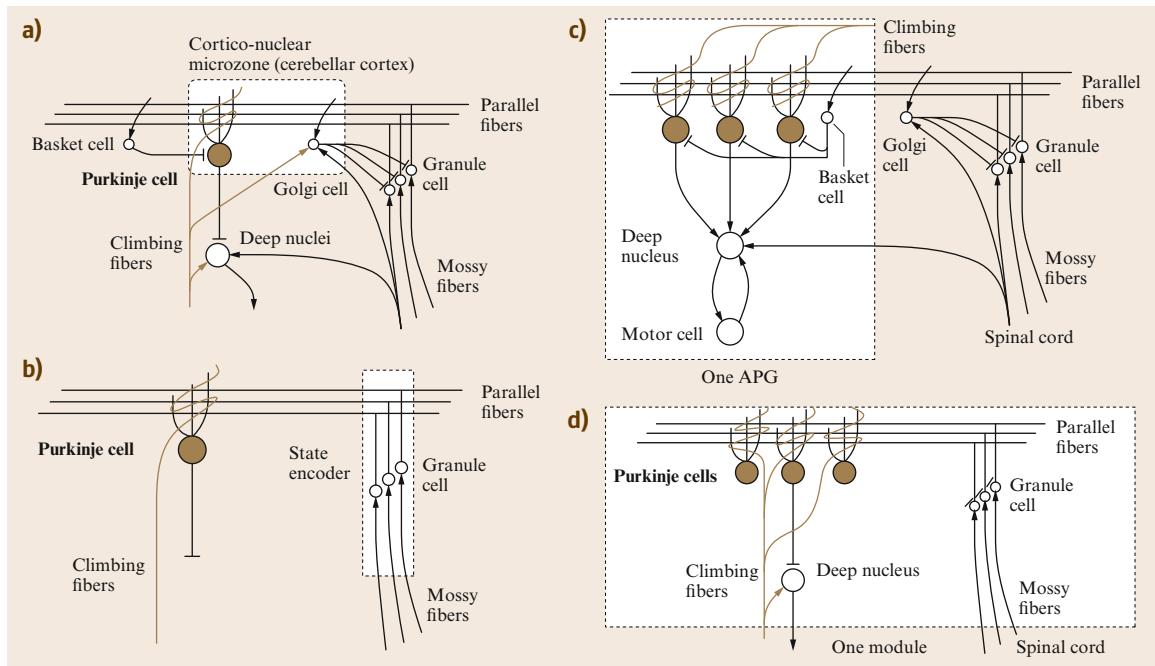**Fig. 77.4** The force/length relationship of muscle (after [77.56])

Fig. 77.5 (a) Major cells in the cerebellum. **(b)** Cells in the Marr–Albus model. The granule cells are state encoders, feeding system state, and sensor data into the PC. PC/PF synapses are adjusted using the Widrow–Hoff rule. The output of the PC are steering signals for the robotic system. **(c)** The APG model, using the same state encoder as in **(b)**. **(d)** The MPFIM model. A single module corresponds to a group of Purkinje cells: predictor, controller, and responsibility estimator. The granule cells generate the necessary basis functions of the original information (after [77.66])

dently, a neural signal controls groups of muscles that perform (a part of) an action. Linear dimension reduction methods [77.69] (e.g., principal component analysis, (PCA), independent component analysis (ICA), or non-negative matrix factorization (NMF)) have been used to establish synergies in EMG data, and this can be used [77.70] to linearly combine single-finger movement to whole-hand movement in EMG space. So, we cannot control single muscles (i. e., coherent groups of muscle fibers) but rather control muscle groups, the linear combination of which can be used to span a decent part of our voluntary movement.

There are currently still open questions as to the nature of movement synergies: how much of the synergies are defined by the biomechanical structure of our muscles and tendons; how much of it is laid out in the spinal cord; and which part of it is learned in the higher movement control regions?

77.3.3 Models of Cerebellar Control

The cerebellum can be divided into two parts: the cortex and the deep nuclei. There are two systems of fibers bringing input to the both the cortex and nuclei: the mossy fibers and the climbing fibers. The only output

from the cerebellar nucleus comes from cells called Purkinje cells, and they project only to the cerebellar nuclei, where their effect is inhibitory. This inhibition sculpts the output of the nuclei which (the effect varies from nucleus to nucleus) may act by modulating activity in the spinal cord, the mid-brain, or the cerebral cortex. We now turn to models that make explicit use of the cellular structure of the *cerebellar* cortex (see *Eccles* et al. [77.71] and *Ito* [77.72], and also Fig. 77.5a). The human cerebellum has $7-14$ million Purkinje cells (PCs), each receiving about $200\,000$ synapses. Mossy fibers (MFs) arise from the spinal cord and brainstem. They synapse onto granule cells and deep cerebellar nuclei. Granule cells have axons which each project up to form a T, with the bars of the T forming the parallel fibers (PFs). Each PF synapses on about 200 PCs. The PCs, which are grouped into microzones, inhibit the deep nuclei. PCs with their target cells in cerebellar nuclei are grouped together in microcomplexes [77.72]. Microcomplexes are defined by a variety of criteria to serve as the units of analysis of cerebellar influence on specific types of motor activity. The climbing fibers (CFs) arise from the inferior olive (IO). Each PC receives synapses from only one CF, but a CF makes about 300 excitatory synapses on each PC that it con-

tacts. This powerful input alone is enough to fire the PC, though most PC firing depends on subtle patterns of PF activity. The cerebellar cortex also contains a variety of inhibitory interneurons. The basket cell is activated by PF afferents and makes inhibitory synapses onto PCs. Golgi cells receive input from PFs, MFs, and CFs, and inhibit granule cells.

The Marr–Albus Model

In the Marr–Albus model (*Marr* [77.73] and *Albus* [77.74]) the cerebellum functions as a classifier of sensory and motor patterns received through the MFs. Only a small fraction of the parallel fibers (PF) are active when a Purkinje cell (PC) fires and thus influence the motor neurons. Both Marr and Albus hypothesized that the error signals for improving PC firing in response to PF, and thus MF input were provided by the climbing fibers (CF), since only one CF affects a given PC. However, Marr hypothesized that CF activity would strengthen the active PF/PC synapses using a Widrow–Hoff learning rule, whereas Albus hypothesized they would weaken them. This is an important example of a case where computational modeling inspired important experimentation. Eventually, *Masao Ito* was able to demonstrate that Albus was correct – the weakening of active synapses is now known to involve a process called long-term depression [77.72]. However, the rule with weakening of synapses is still known as the Marr–Albus model, and remains the reference model for studies of synaptic plasticity of the cerebellar cortex. However, both Marr and Albus viewed each PC as functioning as a perceptron whose job it was to control an elemental movement, contrasting with more plausible models in which PCs serve to modulate the involvement of microcomplexes (which include cells of the deep nuclei) in motor pattern generators (e.g., the APG model described below).

Since the development of the Marr–Albus model several cerebellar models have been introduced in which cerebellar plasticity plays a key role. Limiting our overview to computational models, we will describe:

1. The CMAC (cerebellar model articulation controller).
2. The adjustable pattern generator (APG).
3. The Schweighofer–Arbib model.
4. The multiple paired forward-inverse models [77.75, 76].

The Cerebellar Model Articulation Controller

One of the first well-known computational models of the cerebellum is the CMAC (*Albus* [77.77]; Fig. 77.5b). The algorithm was based on Albus' un-

derstanding of the cerebellum, but it was not proposed as a biologically plausible model. The idea has its origins in the BOXES approach, in which for n variables an n-dimensional hypercube stores function values in a lookup table. BOXES suffers from the curse of dimensionality: if each variable can be discretized into D different steps, the hypercube has to store D^n function values in memory. Albus assumed that the mossy fibers provided discretized function values. If the signal on a mossy fiber is in the receptive field of a particular granule cell, it fires onto a parallel fiber. This mapping of inputs onto binary output variables is often considered to be the generalization mechanism in CMAC. The learning signals are provided by the climbing fibers.

Albus' CMAC can be described in terms of a large set of overlapping, multidimensional receptive fields with finite boundaries. Every input vector falls within the range of some local receptive fields. The response of CMAC to a given input is determined by the average of the responses of the receptive fields excited by that input. Similarly, the training for a given input vector affects only the parameters of the excited receptive fields.

The organization of the receptive fields of a typical Albus CMAC with a two-dimensional input space can be described as follows. The set of overlapping receptive fields is divided into C subsets, commonly referred to as *layers*. Any input vector excites one receptive field from each layer, for a total of C excited receptive fields for any input. The overlap of the receptive fields produces input generalization, while the offset of the adjacent layers of receptive fields produces input quantization. The ratio of the width of each receptive field (input generalization) to the offset between adjacent layers of receptive fields (input quantization) must be equal to C for all dimensions of the input space. This organization of the receptive fields guarantees that only a fixed number, C, of receptive fields is excited by any input.

If a receptive field is excited, its response equals the magnitude of a single adjustable weight specific to that receptive field. The CMAC output is the average of the weights of the excited receptive fields. If nearby points in the input space excite the same receptive fields, they produce the same output value. The output only changes when the input crosses one of the receptive field boundaries. The Albus CMAC thus produces piecewise-constant outputs. Learning takes place as described above.

CMAC neural networks have been applied in various control situations *Miller* [77.78], starting from adaptation of PID (proportional–integral–derivative) control parameters for an industrial robot arm and

hand–eye systems up to biped walking (see, for instance, *Sabourin* and *Bruneau* [77.79]).

The Adjustable Pattern Generator APG

The APG model (*Houk* et al. [77.80]) got its name because the model can generate a burst command with adjustable intensity and duration. The APG is based on the same understanding of the mossy fiber–granule cell–parallel fiber structure as CMAC, using the same state encoder, but has the crucial difference (Fig. 77.2c) that the role of the nuclei is crucial. In the APG model, each nucleus cell is connected to a motor cell in a feedback circuit. Activity in the loop is then modulated by Purkinje cell inhibition, a modeling idea introduced by *Arbib* et al. [77.81].

The learning algorithm determines which of the PF–PC synapses will be updated in order to improve movement generation performance. This is the traditional credit assignment problem: *which* synapse (the structural credit assignment) must be updated based on a response issued *when* (temporal credit assignment). While the former is solved by the CFs, which are considered binary signals, for the latter eligibility traces on the synapses are introduced, serving as memory for recent activity to determine which synapses are eligible for updates. The motivation for the eligibility signal is this: each firing of a PC cell will take some time to affect the animal's movement, and a further delay will occur before the CF can signal an error in the movement in which the PC is involved. Thus the error signal should not affect those PF–PC synapses that are currently active, but should instead act upon those synapses that affected the activity whose error is now being registered.

The Schweighofer–Arbib Model

The Schweighofer–Arbib model was introduced in *Schweighofer* [77.82]. It does not use the CMAC state encoder but tries to copy the anatomy of the cerebellum. All the cells, fibers, and axons in Fig. 77.2a are included. Several assumptions are made:

1. There are two types of mossy fibers, one type reflecting the desired state of the controlled plant and another carrying information on the current state. A mossy fiber diverges into approximately 16 branches.
2. Granule cells have an average of four dendrites, each of which receive input from different mossy fibers through a synaptic structure called the glomerulus.
3. Three Golgi cells synapse on a granule cell through the glomerulus and the strength of their influence depends on the simulated geometric distance between the glomerulus and the Golgi cell.
4. The climbing fiber connection on nuclear cells as well as deep nuclei is neglected.

Learning in this model depends on directed error information given by the climbing fibers from the inferior olive (IO). Here, long-term depression is performed when the IO firing rate provides an error signal for an eligible synapse, while compensatory but slower increases in synaptic strength can occur when no error signal is present. Schweighofer applied the model to explain several acknowledged cerebellar system functions:

1. Saccadic eye movements
2. Two-link limb movement control *Schweighofer* et al. [77.83, 84]
3. Prism adaptation (*Arbib* et al. [77.85]).

Furthermore, control of a simulated human arm was demonstrated.

Multiple Paired Forward–Inverse Models (MPFIM)

Building on a long history of cerebellar modeling, *Wolpert* and *Kawato* [77.86] proposed a functional model of the cerebellum, which uses multiple coupled predictors and controllers that are trained for control, each being responsible for a small state-space region. The MPFIM model is based on the indirect/direct model approach by Kawato, and is also based on the microcomplex theory. We noted earlier that a microzone is a group of PCs, while a microcomplex combines the PCs of a microzone with their target cells in cerebellar nuclei. In MPFIM, a microzone consists of a set of modules controlling the same degree of freedom and is learned by only one particular climbing fiber. The modules in this microzone compete to control this particular synergy. Inside such a module there are three types of PC, which perform the computations of a forward model, an inverse model, or a responsibility predictor, but all receiving the same input. A single internal model i is considered to be a controller that generates a motor command τ_i and a predictor that predicts the current acceleration. Each predictor is a forward model of the controlled system, while each controller contains an inverse model of the system in a region of specialization. The responsibility signal weights the contribution that this model will make to the overall output of the microzone. Indeed, MPFIM further assumes that each microzone contains n internal models of situations occurring in the control task. Model i generates motor command τ_i, and estimates its own responsibility r_i. The feedforward motor command τ_{ff} consists only of the output of the single models adjusted by the sum of responsibility signals: $\tau_{ff} = \sum r_i \tau_i / \sum r_i$.

The PCs are considered to be roughly linear. The MF inputs carry all necessary information including state information, efference copies of the last motor commands, as well as desired states. Granule cells, and eventually the inhibitory interneurons as well, nonlinearly transform the state information to provide a rich set of basis functions through the PFs. A climbing fiber carries a scalar error signal while each Purkinje cell encodes a scalar output – responsibilities, predictions, and controller outputs are all one-dimensional values. MP-FIM has been introduced with different learning methods: its first implementations were done using gradient descent methods; subsequently, expectation maximization (EM) batch-learning, and hidden Markov chain EM learning were applied.

Comparison of the Models

Summing up, we can categorize the cerebellar models CMAC, APG, Schweighofer–Arbib, and MPFIM as follows:

- *State-encoder-driven models*: This kind of model assumes that the granule cells are on–off types of entities that split up the state space. This kind of model is best suited for, e.g., simple function approximation, and suffers strongly from the curse of dimensionality.
- *Cellular-level models*: Obviously, the most realistic simulations would be at the cellular level. Unfortunately, modeling only a few Purkinje cells at realistic conditions is an immense computational challenge, and other relevant neurons are even less well understood. Still, from the biological point of view this kind of model is the most important since it allows obtaining insight into cerebellar function on cellular level. The first steps in this direction were taken by the Schweighofer–Arbib model.
- *Functional models*: From the computer-science point of view, the most interesting models are based on functional understanding of the cells. In this case, we obtain only a basic insight of the functions of the parts and apply it as a crude approximation. This kind of approach is very promising and MP-FIM, with its emphasis on the use of responsibility signals to combine models appropriately, provides an interesting example of this approach.

Proprioceptive feedback is used for adaptation of the motor programs as well as for updating the forward model stored in the cerebellum. However, the Schweighofer–Arbib model is based on the view that the cerebellum offers not so much a total forward model of the skeletomuscular system as a forward model of the difference between the crude model of the skeletomus-cular system available to the motor planning circuits of the cerebral cortex, and the more intricately parameterized forward model of the skeletomuscular system needed to support fast, graceful movements with minimal use of feedback. This hypothesis is reinforced by the fact that cerebellar lesions do not prohibit motion but substantially reduce its quality, since the forward model of the skeletomuscular system is of lesser quality.

77.3.4 Cerebellar Models and Robotics

From the previous discussions, it is clear that a popular view is that the function of the cerebellum within the motor control loop is to represent a forward model of the skeletomuscular system; but how can these models be used in control?

Our assumption is that the cerebellum stores motor primitive relationships, which can be recalled through a certain state (i. e., sensor plus cerebrum-directed goal) input. These motor primitives perform certain coordinated movements (*synergies*) to, e.g., intercept a ball with a tennis racket. A key property of the underlying spine-controlled musculoskeletal system, however, is that *voluntary movement can be easily interpolated* within the control realm of the spinal cord. With this we mean that the combination of two movement primitives that are *nearby* in the relevant sensor domain will lead to a good prediction. *In one possible interpretation, the spinal cord-based control of our muscular system is approximately linear* or *linearized through internal models* [77.87]. It allows the cerebellum to store or recall movements at any level of granularity, and get good enough results in unlearned areas. There are various papers which, in part, confirm this theory (e.g., *Osu* and *Gomi* show the linear relationship between muscle activation and joint stiffness [77.88] or Höppner et al. between grip force and stiffness [77.89]).

Does this understanding of the human control system help robotics? Biological control algorithms are certainly a result of slow feedback loops and the flexibility of the actuators. One may argue that, as robotic systems move towards their biological counterparts, the control approaches can or must do the same. There are many lines of research investigating the former part; Chaps. 11 and 75. It should be noted that the drive principle that is used to move the joints does not necessarily have a major impact on the outer control loop. Whether McKibben muscles, which are intrinsically flexible but bulky (*van der Smagt* et al. [77.90]), low-dynamics polymer linear actuators, or direct-current (DC) motors with spindles and added elastic components are used does not affect the control approach at the cerebellar level, but rather at the motor control level (cf. the spinal

cord level). Of key importance, however, are the resulting dynamic properties of the system, which are, of course, influenced by its actuators. Linearity of the low-level control system, as we find in biology, is a goal to strive for. Yet technical systems can benefit from advanced modeling approaches, and equally good results can be obtained – yet at the cost of more complex sensing, computation, and less generalizability.

77.4 The Role of Mirror Systems

Area F5 (frontal area 5) in the premotor cortex of the macaque contains, among others, neurons which fire when the monkey executes a specific manual action, e.g., one neuron might fire when the monkey performs a precision pinch, another when it executes a power grasp. (In discussing neurorobotics, it seems unnecessary to explain in any detail the areas like F5, AIP (anterior intraparietal sulcus), and STS (superior temporal sulcus) described here – they will function as labels for components of functional systems. To fill in the missing details see, e.g., *Rizzolatti* et al. [77.91, 92].)

77.4.1 Mirror Neurons and the Recognition of Hand Actions

A subset of these neurons, the so-called *mirror neurons*, also discharge when the monkey observes meaningful hand movements made by the experimenter, which are similar to those whose execution is associated with the firing of the neuron. In contrast, the *canonical neurons* are those belonging to the complementary, anatomically segregated subset of grasp-related F5 neurons, which fire when the monkey performs a specific action and also when it sees an object as a possible target of such an action – but do not fire when the monkey sees another monkey or human perform the action. Finally, F5 contains a large population of motor neurons that are active when the monkey grasps an object (either with the hand or mouth) but do not possess any visual response. F5 is clearly a motor area although the details of the muscular activation are abstracted out – F5 neurons can be effector-independent. In contrast, the primary motor cortex (F1) formulates the neural instructions for lower motor areas and motor neurons.

Moreover, macaque mirror neurons encode transitive actions and do not fire when the monkey sees the hand movement unless it can also see the object or, more subtly, if the object is not visible but is appropriately *located* in working memory because it has recently been placed on a surface and has then been obscured by a screen behind which the experimenter is seen to be reaching (*Umiltà* et al. [77.93]). All mirror neurons show visual generalization. They fire when the instrument of the observed action (usually a hand) is large or small, far from or close to the monkey.

They may also fire even when the action instrument has shapes as different as those of a human or monkey hand. Some neurons respond even when the object is grasped by the mouth. When naive monkeys first see small objects grasped with a pair of pliers, mirror neurons do not respond, but after extensive training some precision pinch mirror neurons do show activity, also with this new grasp type [77.94].

Mirror neurons for grasping have also been found in parietal areas of the macaque brain and, recently, it was shown that parietal mirror neurons are sensitive to the context of the observed action being predictive of the outcome as a function of contextual cues – e.g., some grasp-related parietal mirror neurons may fire for a grasp that precedes eating the grasped object, while others fire for a grasp that precedes placing the object in a container (*Fogassi* et al. [77.95]). In practice, the parieto-frontal circuitry seems to encode action execution and simultaneously action recognition by taking into account a large set of potential candidate actions, which are selected on the basis of a range of cues such as vision of the relation of the effector to the object and certain sounds (when relevant for the task). Further, feedback connections (frontal to parietal) are thought to be part of a stimulus selection process that refines the sensory processing by *attending* to stimuli relevant for the ongoing action (*Rizzolatti* et al. [77.52] and recall the discussion in Sect. 77.2.4). Recognition is then supported by the activation of the same circuitry in the absence of overt movement.

We clarify these ideas by briefly presenting the FARS model of the canonical F5 neurons and the MNS model of the F5 mirror neurons. In each case, the F5 neurons function effectively only because of the interaction of F5 with a wide range of other regions. We have stressed (Sect. 77.2.3) the distinction between recognition of the category of an object and recognition of its *affordances*. The parietal area AIP processes visual information to extract *affordances*, in this case properties of the object relevant to grasping it (*Taira* et al. [77.96]). AIP and F5 are reciprocally connected, with AIP being more visual and F5 more motoric.

The Fagg–Arbib–Rizzolatti–Sakata (FARS) model (*Fagg* and *Arbib* [77.97] and Fig. 77.6) embeds F5 canonical neurons in a larger system. The dorsal stream

(which passes through AIP) can only analyze the object as a set of possible affordances, whereas the ventral stream (via the inferotemporal cortex, IT) is able to recognize what the object is. The latter information is passed to the prefrontal cortex (PFC) which can then, on the basis of the current goals of the organism, bias the choice of affordances appropriate to the task at hand. Neuroanatomical data (as analyzed by *Rizzolatti* and *Luppino* [77.98]) suggest that PFC and IT may modulate action selection at the level of the parietal cortex. Figure 77.6 gives a partial view of the FARS model updated to show this modified pathway. The affordance selected by AIP activates F5 neurons to command the appropriate grip once they receive a *go* signal from another region, F6, of the prefrontal cortex. F5 also accepts signals from other PFC areas to respond to working memory and instruction stimuli in choosing among the available affordances. Note that this same pathway could be implicated in tool use, bringing in semantic knowledge as well as perceptual attributes to guide the dorsal system (*Johnson–Frey* [77.99]).

With this, we turn to the *mirror* system. Since grasping a complex object requires careful attention to motion of, e.g., fingertips relative to the object, we hold that the primary evolutionary impetus for the mirror system was to facilitate feedback control of dexterous movement. We now show how *parameters relevant to such feedback* could be crucial in enabling the monkey to associate the visual appearance of what it is doing with the task at hand. The key side-effect will be that this feedback-serving self-recognition is so structured as to also support recognition of the action when performed by others – and it is this *recognition of the*
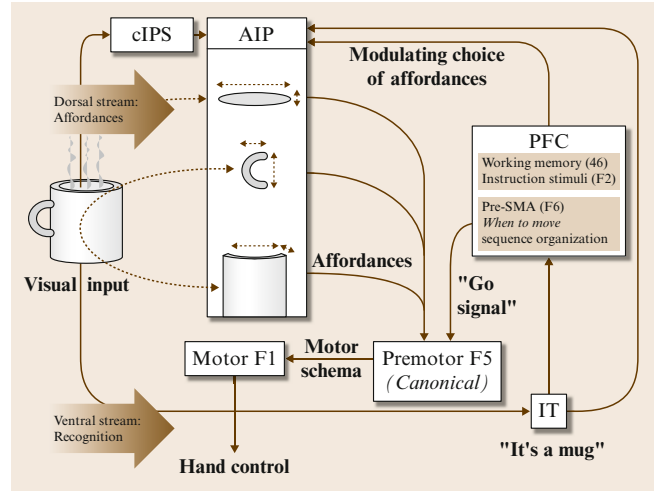

Fig. 77.6 The original FARS diagram (after *Fagg* and *Arbib* [77.42]) is here modified to show PFC acting on AIP rather than F5. The idea is that the prefrontal cortex uses the IT identification of the object, in concert with task analysis and working memory, to help the AIP select the appropriate affordance from its *menu*

actions of others that has created the greatest interest in mirror neurons and systems.

The MNS model of *Oztop* and *Arbib* [77.101] provides some insight into the anatomy while focusing on the learning capacities of mirror neurons. Here, the task is to determine whether the shape of the hand and its trajectory are *on track* to grasp an observed affordance of an object using a known action. The model is organized around the idea that the AIP → F5$_{canonical}$ pathway emphasized in the FARS model (Fig. 77.6) is
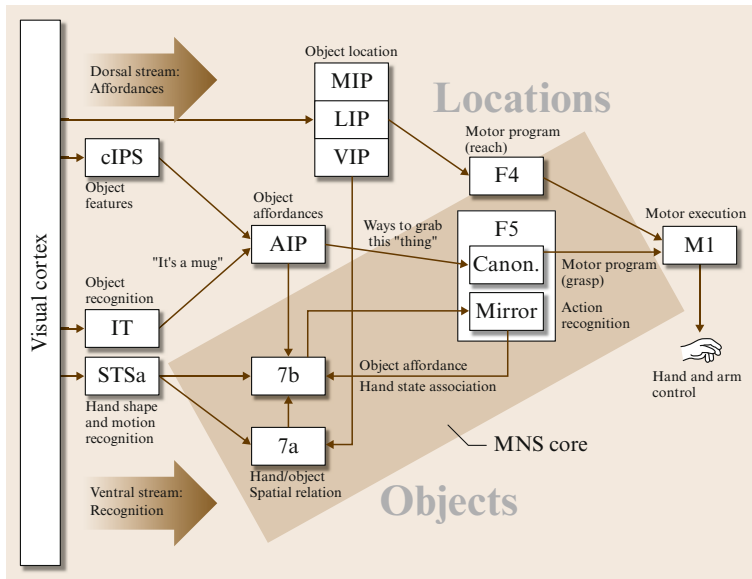
Fig. 77.7 The mirror neuron system (MNS) model (after *Oztop* and *Arbib* [77.100]). Note that this basic mirror system for grasping crucially links the visual process of the STS to the parietal regions (b) and premotor regions (F5), which have been shown to contain mirror neurons for manual actions

complemented by another pathway 7b $\rightarrow$ F5$_{mirror}$. As shown in Fig. 77.7 (middle diagonal), object features are processed by AIP to extract grasp affordances; these are sent on to the canonical neurons of F5 that choose a particular grasp. Recognizing the location of the object (top diagonal) provides parameters to the motor programming area F4 which computes the reach. The information about the reach and the grasp is taken by the motor cortex M1 ($=$ F1) to control the hand and the arm. The rest of the figure provides components that can learn and apply key criteria for activating a mirror neuron, recognizing that the preshape of the observed hand corresponds to the grasp that the mirror neuron encodes and is appropriate to the object, and that the hand is moving on an appropriate trajectory. Making crucial use of input from the superior temporal sulcus (*Perrett* et al. [77.102] and *Carey* et al. [77.103]), schemas at the bottom left recognize the shape of the observed hand and how that hand is moving. Other schemas implement hand–object spatial relation analysis and check how object affordances relate to hand state. Together with F5 canonical neurons, this last schema (in parietal area 7b) provides the input to the F5 mirror neurons.

In the MNS model, the *hand state* was defined as a vector whose components represented the movement of the wrist relative to the location of the object and of the hand shape relative to the affordances of the object. *Oztop* and *Arbib* showed that an artificial neural network corresponding to PF and F5$_{mirror}$ could be trained to recognize the grasp type from the *hand state trajectory*, with correct classification often being achieved well before the hand reached the object, using activity in the F5 canonical neurons that commands a grasp as training signal for recognizing it visually; this basically shows that there is a causal relationship. Crucially, this training prepares the F5 mirror neurons to respond to hand–object relational trajectories even when the hand is of the *other* rather than the *self* because the hand state is based on the view of movement of a hand relative to the object, and thus only *indirectly* on the retinal input of seeing the hand and object, which can differ greatly between observation of self and other. *Bonaiuto* et al. [77.104] have developed MNS2, a new version of the MNS model to address data on audiovisual mirror neurons that respond to the sight and sound of actions with characteristic sounds such as paper tearing and nut cracking *Kohler* et al. [77.93], and on the response of mirror neurons when the target object was recently visible but is currently hidden *Umiltà* et al. [77.93]. Such learning models, and the data they address, make it clear that:

> *mirror neurons are not restricted to recognition of an innate set of actions but can be recruited to rec-*

ognize and encode an expanding repertoire of novel actions.

The discussion of this section avoided any reference to imitation (Sect. 77.4.3). On the other hand, even without considering imitation, mirror neurons provide a new perspective for tackling the problem of robotic perception by incorporating action (and motor information) into a plausible recognition process. The role of the fronto-parietal system in relating affordances, plans, and actions shows the crucial role of motor information and embodiment. We argue that this holds lessons for neurorobotics: the richness of the *motor* system should strongly influence what the robot can learn, proceeding autonomously via a process of exploration of the environment rather than overly relying on the intermediary of logic-like formalisms. When recognition exploits the ability to act, then the breadth of the action space becomes crucially related to the precision, quality, and robustness of the robot's perception.

77.4.2 Computational Models

Roboticists have been fascinated by the discovery of mirror neurons and the purported link to imitation that exists in the human nervous system, for they can help to teach robots new tasks with relative ease. The literature on the topic extends from models of the monkey's (nonimitative) action recognition system (*Oztop* and *Arbib* [77.101]) to models of the putative role of the mirror system in imitation (*Demiris* and *Johnson* [77.105] and *Arbib* et al. [77.106]), and in real and virtual robots (*Schaal* et al. [77.107]). *Oztop* et al. [77.108] propose a taxonomy of the models of the mirror system for recognition and imitation, and it is interesting to note how different the computational approaches that have now been framed as mirror system models are, including recurrent neural networks with parametric bias (*Tani* et al. [77.109]), behavior-based modular networks (*Demiris* and *Johnson* [77.105]), associative memory-based methods (*Kuniyoshi* et al. [77.110]), and the use of multiple direct-inverse models as in the MOSAIC architecture (*Wolpert* et al. [77.111]; cf. the multiple paired forward-inverse models of Sect. 77.3.2).

Following [77.112], we can cast much that is known about the mirror system into a controller-predictor model [77.65, 113] and analyze the resulting model as a Bayesian classifier. As shown by the FARS model, the decision to initiate a particular grasping action is attained by the convergence in area F5 of several factors, including contextual and object-related information; similarly many factors affect the recognition of an action. All this depends on learning both direct (from decision to executed action) and inverse models

(from observation of an action to activation of a motor command that could yield it). Similar procedures are well known in the computational motor control literature [77.114, 115]. Learning of the affordances of objects with respect to grasping can also be achieved autonomously by learning from the consequences of applying many different actions to different parts of different objects.

However, how is the decision made to classify an observed behavior as an instance of one action or another? Many comparisons could be performed in parallel with the models for one action to become predominantly activated. There are plausible implementations of this mechanism using a gating network [77.105, 116]. A gating network learns to partition an input space into regions; for each region a different model can be applied or a set of models can be combined through an appropriate weight function. The design of the gating network can encourage collaboration between models (e.g., linear combination of models) or competition (choosing only one model rather than a combination). Reference [77.117] offers a similar approach to the estimation of the mental states of the observed actor, using some additional circuitry involving the frontal cortex.

On the other hand, if we take the Bayesian view of the predictor-controller formulation, then affordances are simply the priors in the action recognition process where the evidence is conveyed by the visual information of the hand, providing the data for finding the posterior probabilities as mirror neuron-like responses which automatically activate for the most probable observed action. Recall that the presence of a goal (at least in working memory) is needed to elicit mirror neuron responses in the macaque. We believe it is also particularly important during the ontogenesis of the human mirror system. For example, [77.118] has shown that even at 9 months of age, infants recognized an action as being novel if it was directed toward a novel object rather than just having a different kinematics – showing that the goal is more fundamental than the enacted trajectory. Similarly, if one sees someone drinking from a coffee mug then one can hypothesize that a particular action (that one already knows in motor terms) is used to obtain that particular effect. The association between the canonical response (object-action) and the mirror one (including vision) is made when the observed consequences (or goal) are recognized as similar in the two cases. Similarity can be evaluated following criteria ranging from kinematic to social consequences.

In a similar experiment *Lopes* et al. [77.119] compared action recognition performance (a) when using the output of an inverse visuo-motor model and thus employing motor features to aid classification during the training phase, and (b) when only visual data were available for recognition. Overall, their interpretation of the results is that by mapping in motor space through inverse model mapping, they allow the classifier to choose features that are much better suited for performing optimally with respect to the task of recognizing actions, which in turn facilitates generalization. The same is not true when recognition is performed purely in visual space using generic visual features, since a given action is viewed from different viewpoints. One may compare this to the viewpoint-invariant hand state adopted in the MNS model – which has the weakness of being built in rather than emerging from training.

Along the same line, the work of *Gijsberts* et al. [77.120] included motorically-derived affordance information, which was recorded using a data-glove-based system and a set of cameras. In this case though, motor information was not much for action recognition but rather used to *simulate* the response of F5's canonical neurons by generating discrete grasping types from the time-varying set of postures recorded with the data glove. After training the original motor information is removed and only reconstructed using an inverse model. Furthermore, this motoric information was combined with a simulation of the brain ventral pathway which extracts pictorial features from images (e.g., SIFT (scale-invariant feature transform), H-Max). The dorsal and ventral features were combined through a special kernel function in a simple least squares classifier, showing a significant improvement at recognizing objects in comparison to a purely visual classification. A machine learning framework to address the question of learning from multimodal signals (some of which can even be intermittent) is presented in [77.121].

We can speculate that this computational advantage (better recognition rates) makes the presence of mixed sensory and motor information compelling in the brain (i.e., the fronto-parietal system); this may not necessarily lead to mirror neurons although it seems plausible that any clear advantage of using information at best is eventually selected during evolution. These experiments, using robots, simulations, and computational arguments can thus explain the whys of certain brain structures and mechanisms.

77.4.3 Mirror Neurons and Imitation

Fitzpatrick and *Metta* [77.122] also addressed the question of what is further required for interpreting observed actions. Whereas in observing its own actions, the robot identifies them from the effects on the objects, later it could backtrack and derive the type of action needed to replicate a certain observed effect on a given object.

Therefore, imitation can be framed into the identification of a common goal between the observed action and various possible actions in the *motor repertoire* of the robot. In [77.122] the robot used the same visual processing algorithms both in observing its own hand and the hand of a person (although they were different in appearance). One might argue that observation alone can be used for learning, never relying on active exploration of objects and actions. This is possibly true to the extent that passive vision is reliable and action is not required. The advantage of the active approach, at least for the robot, is that it allows controlling the amount of information impinging on the visual sensors by, for instance, controlling the speed and type of action. This strategy might be especially useful given the limitations of artificial perceptual systems. Thus, observations can be converted into interpreted actions. The action whose effects are closest to the observed consequences on the object (which we might translate into the goal of the action) is selected as the most plausible interpretation given the observation. Most importantly, the interpretation reduces to the interpretation of the *simple* kinematics of the goal and consequences of the action rather than to understanding the *complex* kinematics of the human manipulator. The robot understands only to the extent it has learned to act. One might note that a more refined model should probably include visual cues from the appearance of the manipulator into the interpretation process. Indeed, the *hand state* that was central to the *Oztop–Arbib* model was based on an object-centered view of the hand's trajectory in a coordinate frame based on the object's affordances. The last question to address is whether a robot can imitate the *goal* of the action. The step is indeed small, since most of the complexity is actually in interpreting observations. Imitation can be generated by replicating the latest observed human movement with respect to the object utilizing one of the many approximation methods for motion generation such as, e.g., a mixture of Gaussians [77.123], dynamic motion primitives [77.124], or reinforcement learning [77.125]. More generally, following the work of *Schaal* et al. [77.107] and *Oztop* et al. [77.108] we can propose a set of schemas required to produce imitation:

- Determining what to imitate, inferring the goal of the demonstrator
- Establishing a metric for imitation (correspondence; see *Nehaniv* [77.126])
- Map between dissimilar bodies (mapping).
- Imitating behavior formation.

These are also discussed in greater detail by *Nehaniv* and *Dautenhahn* [77.127]. In practice, computa-

tional and robotic implementations have tackled these problems with different approaches and emphasizing different parts or specific subproblems of the whole, for example, in the work of *Demiris* and *Hayes* [77.128], the rehearsal of the various actions (akin to the aforementioned theory of motor perception) was used to generate hypotheses to be compared with the actual sensory input. It is then remarkable how more recently a modified approach of this paradigm has been used in comparison with real human transcranial magnetic stimulation (TMS) data.

Ito et al. [77.129] (not Masao Ito of cerebellar fame) took a dynamical systems approach using a recurrent neural network with parametric bias (RNNPB) to teach a humanoid robot to manipulate certain objects. In this approach the parametric bias (PB) encodes (tags) certain sensorimotor trajectories. Once learning is complete, the neural network can be used either to recall a given trajectory by setting the PB externally or provide input for the sensory data only and observe the PB vector that would represent in that case the recognition of the situation on the basis of the sensory input only (no motor information available). It is relatively easy to interpret these two situations as the motor generation and the observation in a mirror neurons model.

The problem of building useful mappings between dissimilar bodies (consider a human imitating a bird's flapping wings) was tackled by *Nehaniv* and *Dautenhahn* [77.127] where an algebraic framework for imitation is described and the correspondence problem formally addressed. Any system implementing imitation should clearly provide a mapping between either dissimilar bodies or even in the case of similar bodies when either the kinematics or dynamics is different depending on the context of the imitative action.

Sauser and *Billard* [77.130] modeled the *ideomotor principle*, according to which observing the behavior of others influences our own performances. The ideomotor principle points directly to one of the core issues of the mirror system, that is, the fact that watching somebody else's actions changes something in the activation of the observer, thus facilitating certain neural pathways. The work in question also gives a model implemented in terms of neural fields (see *Sauser* and *Billard* [77.130] for details) and tries to explain the imitative cortical pathways and the behavior formation.

77.4.4 Mirror System and Speech

Already in the 1960s *Liberman* et al. [77.131] started to discuss the possible links between production and perception in speech: in other words the contribution of articulation into the perception of utterances. Later he commented [77.132]:

A result in all cases is that there is not, first, a cognitive representation of the proximal pattern that is modality-general, followed by a translation to a particular distal property; rather, perception of the distal property is immediate, which is to say that the module has done all the hard work.

Liberman argued that there is no such a thing as a modality-aspecific representation which then becomes speech as an effect of a translation to a specific set of articulators (the vocal apparatus in this case), rather, he claimed that perception of speech is immediate and effected by the same speech module (the same that generates speech); speech remains a motor fact. Lately, theories of the motor involvement in speech perception have gained credit because of the discovery of the mirror neurons. It has been postulated that the mirror system in humans *controls* jointly speech production and perception, whereby the actions in speech are the articulation of appropriate segments of the utterances [77.133].

We recall this line of reasoning in the following [77.133], that is:

- Mirror neurons (or a mirror system) exist in humans [77.134].
- The human mirror system is identified in Broca's area, a cytoarchitectonical homolog of area F5 in the macaque's brain.
- Speech articulation is coded/controlled in/by the areas of the human mirror system (Broca's) [77.135].
- The recognition of the intention of the speaker by the listener owing to a mirror mechanism leads to the first seed of true communication (via, e.g., oro-facial gestures) [77.133].
- The combinatorial properties of F5/Broca and the precise control of the effectors are needed to generate speech (the evolutionarily older animal calls are too stereotyped to grant this flexibility that eventually leads to speech proper) [77.136].

To establish that this is the case, however, more empirical evidence is required. Recently, two experiments improved the plausibility of the mirror neurons theory of speech perception. In a first TMS experiment, *Fadiga* and colleagues [77.137] established that MEPs (motor evoked potentials) in the tongue muscles directly correlate with high specificity to the perception of particular sounds (these were *rr* and *ff* in Italian). The listener was delivered TMS (single pulse) and the observed MEPs correlated in amplitude with the different use of the tongue muscles for the pronunciation of either the *rr* or *ff* sounds (*rr* in Italian requires a strong mobilization of the tongue). Albeit convincing, this experiment leaves

open the question of specificity, since it can still be the case of a diffuse/generic activation of Broca's area.

A second experiment also by *Fadiga* et al. [77.137] was designed to set the issue. In this case, the TMS was delivered to the primary motor cortex with the aim of establishing a specific motor involvement into the perception of different sounds/phones. Two areas were individuated in the primary motor cortex as responsible to the lip and tongue movement, respectively (e.g., *p/b* sounds versus t/d). The data show a double dissociation pattern, that is, when the lip motor area is stimulated there is a decrease of the reaction times (RTs) of the subject in perceiving the *p/b* (labial sounds) and vice-versa an increase for the perception of the t/d (dental sounds). The opposite happens when the tongue motor area is stimulated. This experiment clearly relates a very specific (small) region of the primary motor cortex with the perception of certain specific (and related) sounds.

Clearly, this is only part of the story; to complicate matters, for example, the semantic content of words related to actions (e.g., kick, pick, lick) activate both motor and pre-motor brain areas somatotopically. Object features, odors, etc., instead have been shown to generate responses in the corresponding cortices. For a review of these and other results, see [77.136].

Theories and models such as the perception for action control theory (PACT) [77.138] take a more moderate interpretation by including both a motor component and *perceptual shaping*, that is, the filtering of certain linguistic combinations because of purely perceptual characteristics (e.g., separation of vowel formants). In PACT, it is hypothesized that the motor system is activated more in *adverse conditions*, while it is perhaps under-threshold for normal speech understanding in good signal-to-noise conditions.

Indeed, we can recognize speech in a foreign accent, and recognizing what is being said can then be decoupled from being able to articulate how it is being said – but both possibilities are available. This has led to a new view of the integration of mirror systems with other systems [77.139] which downplays the motor theory of speech perception while preserving many other features of the mirror system hypothesis of *Rizzolatti* and *Arbib* [77.133].

Armed with these results *Castellini* and colleagues [77.140] conducted a computational experiment that mimic some of the TMS results of *D'Ausilio* et al. [77.141]. All processing employed a database of synchronized recordings of Italian speakers with acoustic, articulograph, camera, ultrasound, and electroglottograph data [77.142]. For the experiments, only the articulograph and electroglottograph signals were used together with speech sound. These identify the position of the tongue and teeth versus the lips in
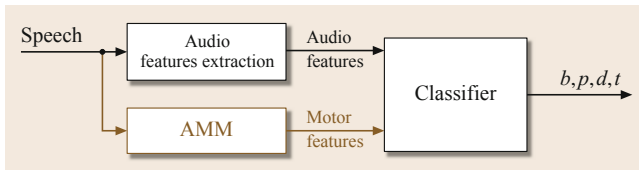
Fig. 77.8 Conceptual schema of the classifiers used in the experiments
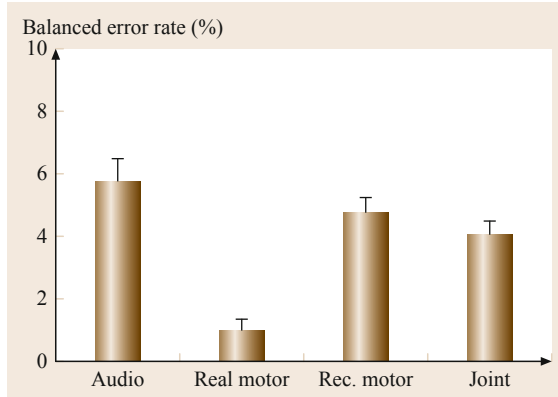

Fig. 77.9 Baseline experiment comparing the performance of acoustic versus motor data (or jointly acoustic plus motor data) in classifying b/p versus d/t

real time (200 Hz) in addition to the activation of the vocal folds (voicing signal). The conceptual schema of all experiments and learning follows some previous work as by *Metta* et al. [77.112], and which as is shown in Fig. 77.8. In particular, acoustic data are mapped into motoric features and these are used for classifying phones. Similarly to the PACT model, it was found useful to incorporate also a purely acoustic classifier. Acoustic features were the standard Mel cepstral coefficients with similar parameters and frequency

bands of conventional automatic speech recognition (ASR).

The mapping from acoustic to motor data was performed using either an artificial neural network or support vector machine for regression with indistinguishable results. The classifier was always a support vector machine with Gaussian kernel and parameters optimized through grid search.

In order to compare it with the TMS experiments, phones were divided into two classes, the b/p and t/d, respectively, as representing the bi-labial and dental (movement of the tongue toward the teeth) phones. Fivefold cross-validation was employed on all results by either random splits of the data or by selecting data from various participants (e.g., training on $1-5$ participants, testing on $1-5$ participants). Gaussian white noise was added to the stimuli (at increasing levels from 0 to 150%) to replicate the conditions of the TMS.

Figure 77.9 shows these results. The baseline experiment shows an improved performance where either the real motoric or jointly motoric and acoustic features are used. The comparison of *audio* versus *joint* features is statistically significant ($p < 0.01$) and verifies the claim as no new information is added to the system when the reconstructed motor features are employed. Motor features are *reconstructed* by the audio-motor map (AMM) of Fig. 77.8 and replicates previous results obtained in the classification of hand gestures [77.119] or handwriting characters [77.143].

A second experiment from the same work of *Castellini* and colleagues [77.140] shows the behavior of the same system in various conditions of increased difficulty ranging from running classification on speakers not included in the training sets to co-articulation. Figure 77.10 shows a number of variants where N vs M indicates N speakers for training versus M speakers for testing given the size of the database (6 speakers). Ex-
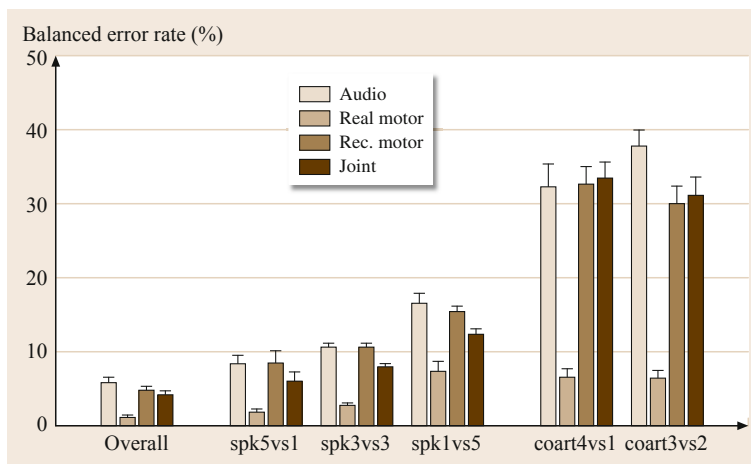

Fig. 77.10 Comparison across various conditions. In all cases apart from coart4vs1 the use of motor features improve classification with statistical significance ($p < 0.01$)
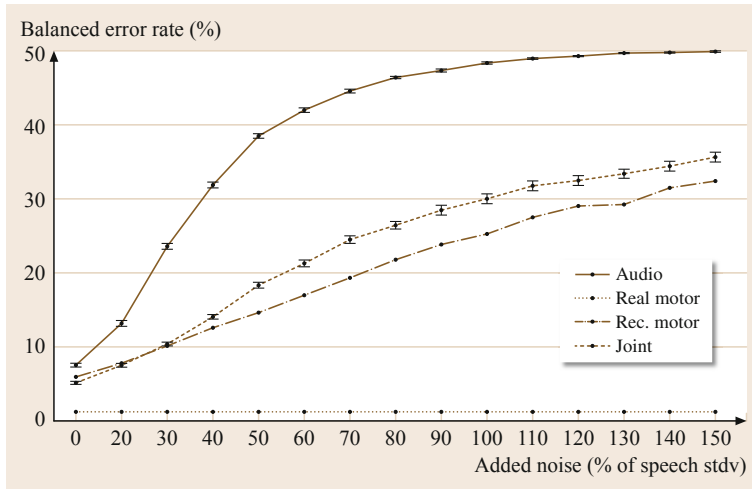
Fig. 77.11 Comparison of acoustic versus motor features under increasing level of added Gaussian white noise for the same classifier of the previous experiments

periments with co-articulation were conducted also on five speakers, albeit the number of identifiable examples in the database was smaller.

In a final experiment, the classifier was tested on acoustic data corrupted by Gaussian white noise. Results show a consistent improvement with the motor information gain increasing with the increase of the noise level (up to 150% of the speech standard deviation): Fig. 77.11.

More recently a full phone classifier was built using similar principles [77.144] together with a combination of deep belief networks (DBNs) and more standard hidden Markov models (HMMs). The results show improvement with respect to the state of the art, continuing the long tradition of neurorobotics and bringing models very close to concrete applications on robots that bear resemblance to the exquisite human performance in speech recognition in noisy environments.

77.5 Conclusion and Further Reading

As the foregoing makes clear, robotics has much to learn from neuroscience and much to teach neuroscience. Neurorobotics can learn from the ways in which the brains and bodies of different creatures adapt to diverse ecological niches – as computational neuroethology helps us understand how the brain of a creature has evolved to serve *action-oriented perception*, and the attendant processes of learning, memory, planning, and social interaction.

We have sampled the *design* of just a few subsystems (both functional and structural) in just a few animals – optic flow in the bee, approach, escape, and barrier avoidance in frogs and toads, and navigation in the rat, as well as the control of eye movements in visual attention, the role of the mammalian cerebellum in handling the nonlinearities and time delays of flexible motor systems, and the mirror systems of primates in action recognition and of humans in imitation. There are many more creatures with lessons to offer the roboticist than we can sample here.

Moreover, if we just confine attention to the brains of humans, this chapter has mentioned at least 7a,

7b, AIP, lateral, medial and ventral intraparietal sulcus (LIP, MIP and VIP), area 46, basal ganglia, caudoputamen, cerebellum, F2, F4, F5, hippocampus, hypothalamus, inferotemporal cortex, motor cortex, nucleus accumbens, parietal cortex, prefrontal cortex, premotor cortex, pre-SMA (F6), spinal cord, STS, and – and it is clear that there are many more details to be understood for each region, and many more regions whose interactions hold lessons for roboticists. We say this not to depress the reader, but rather to encourage further exploration of the literature of computational neuroscience and to note that the exchange with neurorobotics proceeds both ways: neuroscience can inspire novel robotic designs; conversely, robots can be used to test whether brain models still work when they make the transition from disembodied computer simulation to meeting the challenge of guiding the interactions of a physically embodied system with the complexities of its environment.

Nonetheless, a thorough study of the spinal cord and its effect on muscle behavior is where a roboticist, who is interested in replicating some of the functionality of vertebrate movement, may want to start looking.

77.5.1 Further Reading

- *Arbib* (2006) [77.145]: This volume provides 16 articles on the mirror system, written by diverse experts. Of particular relevance to this chapter are articles on dynamical systems: brain, body and imitation; attention and the minimal subscene; the development of grasping and the mirror system; and development of goal-directed imitation, object manipulation, and language in humans and robots.
- *Bell* (1996) [77.146]: This somewhat older BBS special issue provides what was, back then, a rather definitive number of articles on the cerebellum, including an overview of models in a paper by *Houk* et al.
- *van der Smagt* and *Bullock* (2002) [77.147]: This special issue is focused on the application of cere-

bellar and other models to robotics tasks, and lists some successful and – between the lines – more unsuccessful applications thereof.

- *Gallese* et al. (1996) [77.148]: This paper provides a detailed account of the neurophysiological evidence for mirror neurons. It is good reading to get the real data unbiased from further interpretation on the role of mirror neurons and it is complete and accurate. Although it is a technical paper it is easy to read also for a general audience.
- *Fadiga* et al. 2002 [77.149]: This work extends the mirror system concept with an interesting perspective on its role into language. This paper is interesting reading by providing evidence in humans (the other references above are about monkey experiments). In this case, it has been shown that speech listening facilitates the activation of tongue muscles which match the specific phoneme being listened to.

References

77.1   W. G. Walter: *The Living Brain* (Duckworth, London 1953), reprinted by Pelican Books, Harmondsworth, 1961

77.2   V. Braitenberg: *Vehicles: Experiments in Synthetic Psychology* (Bradford Books/The MIT, Cambridge 1984)

77.3   I. Segev, M. London: Dendritic processing. In: *The Handbook of Brain Theory and Neural Networks*, 2nd edn., ed. by M.A. Arbib (Bradford Books/MIT Press, Cambridge 2003) pp. 324–332

77.4   Y. Fregnac: Hebbian synaptic plasticity. In: *The Handbook of Brain Theory and Neural Networks*, 2nd edn., ed. by M.A. Arbib (Bradford Books/MIT Press, Cambridge 2003) pp. 515–522

77.5   W. Reichardt: Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In: *Sensory Communication*, ed. by W.A. Rosenblith (MIT Press/Wiley, New York, London 1961) pp. 303–317

77.6   A. Borst, M. Dickinson: Visual course control in flies. In: *The Handbook of Brain Theory and Neural Networks*, 2nd edn., ed. by M.A. Arbib (Bradford Books/MIT Press, Cambridge 2003) pp. 1205–1210

77.7   P. van der Smagt, F. Groen: Visual feedback in motion. In: *Neural Systems for Robotics*, ed. by O. Omidvar, P. van der Smagt (Morgan Kaufmann, San Francisco 1997) pp. 37–73

77.8   P. van der Smagt: Teaching a robot to see how it moves. In: *Neural Network Perspectives on Cognition and Adaptive Robotics*, ed. by A. Browne (Institute of Physics Publishing, Bristol 1997) pp. 195–219

77.9   S.C. Liu, A. Usseglio-Viretta: Fly-like visuomotor responses of a robot using aVLSI motion-sensitive chips, Biol. Cybern. **85**(6), 449–457 (2001)

77.10   F. Ruffier, S. Viollet, S. Amic, N. Franceschini: Bio-inspired optical flow circuits for the visual guidance of micro air vehicles, Int. Symp. Circuits Syst. (ISCAS) 2003, Vol. 3 (2003)

77.11   M.B. Reiser, M.H. Dickinson: A test bed for insect-inspired robotic control, Philos. Trans. Math. Phys. Eng. Sci. **361**(1811), 2267–2285 (2003)

77.12   M. V. Srinivasan, S. W. Zhang, M. Altwein, J. TAUTZ: Honeybee navigation: Nature and calibration of the odometer, Science **287**, 851–853 (2000)

77.13   S. Funke: *Virtual atlas of the honeybee brain* (see Univ. Berlin, Institute of Biology – Neurobiology, Berlin 2015), http://www.neurobiologie.fu-berlin.de/beebrain/Bee/VRML/SnapshotCosmoall.jpg

77.14   A. Barron, M.V. Srinivasan: Visual regulation of ground speed and headwind compensation in freely flying honey bees, *Apis mellifera L*, J. Exp. Bio. **209**(5), 978–984 (2006)

77.15   M.V. Srinivasan, S. Zhang, J.S. Chahl: Landing strategies in honeybees, and possible applications to autonomous airborne vehicles, Biol. Bull. **200**(2), 216–221 (2001)

77.16   T. Vladusich, J.M. Hemmi, M.V. Srinivasan, J. Zeil: Interactions of visual odometry and landmark guidance during food search in honeybees, J. Exp. Biol. **208**, 4123–4135 (2005)

77.17   M.V. Srinivasan, S.W. Zhang: Visual control of honeybee flight. In: *Orientation and Communication in Arthropods*, Experientia Supplementum, Vol. 84, ed. by M. Lehres (Birkhäuser, Basel 1997) pp. 95–113

77.18   B. Webb: Can robots make good models of biological behavior?, Behav. Brain Sci. **24**, 1033–1094 (2001)

77.19 J.Y. Lettvin, H. Maturana, W.S. McCulloch, W.H. Pitts: What the frog's eye tells the frog brain, Proc. IRE **47**, 1940–1951 (1959)

77.20 D. Ingle: Visual releasers of prey catching behavior in frogs and toads, Brain Behav. Evol. **1**, 500–518 (1968)

77.21 R.L. Didday: A model of visuomotor mechanisms in the frog optic tectum, Math. Biosci. **30**, 169–180 (1976)

77.22 M.A. Arbib: Levels of modeling of visually guided behavior, Behav. Brain Sci. **10**, 407–465 (1987)

77.23 M.A. Arbib: Visuomotor coordination: Neural models and perceptual robotics. In: *Visuomotor Coordination: Amphibians, Comparisons, Models, and Robots*, 2nd edn., ed. by J.P. Ewert, M.A. Arbib (Plenum, New York 1989) pp. 121–171

77.24 T. Collett: Do toads plan routes? A study of detour behavior of *B. viridis*, J. Comp. Physiol. **146**, 261–271 (1982)

77.25 M.A. Arbib, D.H. House: Depth and detours: An essay on visually guided behavior. In: *Vision, Brain and Cooperative Computation*, ed. by M.A. Arbib, A.R. Hanson (Bradford Books/MIT Press, Cambridge 1987) pp. 129–163

77.26 F.J. Corbacho, M.A. Arbib: Learning to detour, Adapt. Behav. **4**, 419–468 (1995)

77.27 A. Cobas, M.A. Arbib: Prey-catching and predator-avoidance in frog and toad: Defining the schemas, J. Theor. Biol. **157**, 271–304 (1992)

77.28 R.C. Arkin: Motor schema-based mobile robot navigation, Int. J. Robot. Res. **8**, 92–112 (1989)

77.29 O. Khatib: Real-time obstacle avoidance for manipulators and mobile robots, Int. J. Robot. Res. **5**, 90–98 (1986)

77.30 B.H. Krogh, C.E. Thorpe: Integrated path planning and dynamic steering control for autonomous vehicles, Proc. IEEE Int. Conf. Robot. Autom. San Francisco (1986) pp. 1664–1669

77.31 R.A. Brooks, C.L. Breazeal, M. Marjanović, B. Scassellati: The COG project: Building a humanoid robot, Lect. Notes Comput. Sci. **1562**, 52–87 (1999)

77.32 R.C. Arkin: *Behavior-Based Robotics* (MIT Press, Cambridge 1998)

77.33 R.C. Arkin, M. Fujita, T. Takagi, R. Hasegawa: An ethological and emotional basis for human-robot interaction, Robot. Auton. Syst. **42**(3/4), 191–201 (2003)

77.34 P. Dean, P. Redgrave, G.W.M. Westby: Event or emergency? Two response systems in the mammalian superior colliculus, Trends Neurosci. **12**, 138–147 (1989)

77.35 B.E. Stein, M.A. Meredith: *The Merging of the Senses* (MIT Press, Cambridge 1993)

77.36 T. Strosslin, C. Krebser, A. Arleo, W. Gerstner: Combining multimodal sensory input for spatial learning, artificial neural networks – ICANN 2002, Lect. Notes Comput. Sci. **2415**, 87–92 (2002)

77.37 J. O'Keefe, L. Nadel: *The Hippocampus as a Cognitive Map* (Clarendon, Oxford 1978)

77.38 V. Braitenberg: Taxis, kinesis, decussation, Progr. Brain Res. **17**, 210–222 (1965)

77.39 A. Guazzelli, F.J. Corbacho, M. Bota, M.A. Arbib: Affordances, motivation, and the world graph theory, Adapt. Behav. **6**, 435–471 (1998)

77.40 J.J. Gibson: *The Senses Considered as Perceptual Systems* (Allen and Unwin, London 1966)

77.41 I. Lieblich, M.A. Arbib: Multiple representations of space underlying behavior, Behav. Brain Sci. **5**, 627–659 (1982)

77.42 B. Girard, D. Filliat, J.A. Meyer, A. Berthoz, A. Guillot: Integration of navigation and action selection functionalities in a computational model of cortico-basal-ganglia-thalamo-cortical loops, Adapt. Behav. **13**(2), 115–130 (2005)

77.43 J.A. Meyer, A. Guillot, B. Girard, M. Khamassi, P. Pirim, A. Berthoz: The Psikharpax project: Towards building an artificial rat, Robot. Auton. Syst. **50**(4), 211–223 (2005)

77.44 G. Deco, E.T. Rolls: Attention and working memory: A dynamical model of neuronal activity in the prefrontal cortex, Eur. J. Neurosci. **18**(8), 2374–2390 (2003)

77.45 S.B. Choi, S.W. Ban, M. Lee: Biologically motivated visual attention system using bottom-up saliency map and top-down inhibition, Neural Inf. Proc. – Lett. Rev. **2**(1), 19–25 (2004)

77.46 L. Itti, C. Koch: A saliency-based search mechanism for overt and covert shifts of visual attention, Vis. Res. **40**, 1489–1506 (2000)

77.47 J.M. Wolfe: Guided search 2.0: A revised model of visual search, Psychon. Bull. Rev. **1**, 202–238 (1994)

77.48 A. Yarbus: *Eye Movements and Vision* (Plenum, New York 1967)

77.49 V. Navalpakkam, L. Itti: Modeling the influence of task on attention, Vis. Res. **45**, 205–231 (2005)

77.50 F. Orabona, G. Metta, G. Sandini: Object-based visual attention: A model for a behaving robot, Conf. Comput. Vis. Pattern Recogn. (2005) pp. 89–89

77.51 G. Sandini, V. Tagliasco: An anthropomorphic retina-like structure for scene analysis, Comput. Vis. Gr. Image Proc. **14**(3), 365–372 (1980)

77.52 G. Rizzolatti, L. Riggio, I. Dascola, C. Umiltá: Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention, Neuropsychologia **25**, 31–40 (1987)

77.53 J.R. Flanagan, R.S. Johansson: Action plans used in action observation, Nature **424**, 769–771 (2003)

77.54 J.R. Flanagan, P. Vetter, R.S. Johansson, D.M. Wolpert: Prediction precedes control in motor learning, Curr. Biol. **13**, 146–150 (2003)

77.55 J.M. Mataric, M. Pomplun: Fixation behavior in observation and imitation of human movement, Brain Res. Cogn. Brain Res. **7**, 191–202 (1998)

77.56 E. Kandel, J. Schwartz, T. Jessell: *Principles of Neural Science*, 5th edn. (McGraw-Hill, Columbus 2000)

77.57 J. Fallon, R. Carr, D. Morgan: Stochastic resonance in muscle receptors, J. Neurophysiol. **91**, 2429–2436 (2004)

77.58 B. Edin: Quantitative analyses of dynamic strain sensitivity in human skin mechanoreceptors, J. Neurophysiol. **92**(6), 3233–3243 (2004)

77.59 K.O. Johnson: Closing in on the neural mechanisms of finger joint angle sense. Focus on quantitative analysis of dynamic strain sensitivity in human skin mechanoreceptors, J. Neurophysiol. **92**(6), 3167–3168 (2004)

77.60 B. Edin: Cutaneous afferents provide information about knee joint movements in humans, J. Physiol. **531**(1), 289–297 (2001)

77.61 B. Edin, J. Westlin: Independent control of human finger–tip forces at individual digits using precision lifting, J. Physiol. **450**, 547–567 (1992)

77.62 K. Doya: Complementary roles of basal ganglia and cerebellum in learning and motor control, Curr. Opin. Neurobiol. **10**(6), 732–739 (2000)

77.63 L.M. Bjursten, K. Norrsell, U. Norrsell: Behavioural repertory of cats without cerebral cortex from infancy, Exp. Brain Res. **25**(2), 115–130 (1976)

77.64 G. Holmes: The cerebellum of man, Brain **62**, 1–30 (1939)

77.65 R.C. Miall, D.J. Weir, D.M. Wolpert, J.F. Stein: Is the cerebellum a Smith predictor?, J. Motor Behav. **25**, 203–216 (1993)

77.66 J. Peters, P. van der Smagt: Searching a scalable approach to cerebellar based control, Appl. Intell. **17**, 11–33 (2002)

77.67 E.J. Nijhof, E. Kouwenhoven: Simulation of multijoint arm movements. In: *Biomechanics and Neural Control of Posture and Movement*, ed. by J.M. Winters, P.E. Crago (Springer, New York 2002) pp. 363–372

77.68 N. Bernstein: *The Coordination and Regulation of Movements* (Pergamon, Oxford 1967)

77.69 M.C. Tresch, V.C.K. Cheung, A. d'Avella: Matrix factorization algorithms for the identification of muscle synergies: Evaluation on simulated and experimental data sets, J. Neurophysiol. **95**(4), 2199–2212 (2006)

77.70 C. Castellini, P. van der Smagt: Evidence of muscle synergies during human grasping, Biol. Cybern. **107**(2), 233–245 (2013)

77.71 J.C. Eccles, M. Ito, J. Szentágothai: *The Cerebellum as a Neuronal Machine* (Springer, New York 1967)

77.72 M. Ito: *The Cerebellum and Neural Control* (Raven, New York 1984)

77.73 D.A. Marr: A theory of cerebellar cortex, J. Physiol. **202**, 437–470 (1969)

77.74 J.S. Albus: A theory of cerebellar function, Math. Biosci. **10**, 25–61 (1971)

77.75 P. van der Smagt: Cerebellar control of robot arms, Connect. Sci. **10**, 301–320 (1998)

77.76 P. van der Smagt: Benchmarking cerebellar control, Robot. Auton. Syst. **32**, 237–251 (2000)

77.77 J.S. Albus: Data storage in the cerebellar model articulation controller (CMAC), J. Dyn. Syst. Meas. Control ASME **3**, 228–233 (1975)

77.78 W.T. Miller: Real–time application of neural networks for sensor–based control of robots with vision, IEEE Trans. Syst. Man Cybern. **19**, 825–831 (1994)

77.79 C. Sabourin, O. Bruneau: Robustness of the dynamic walk of a biped robot subjected to disturbing external forces by using CMAC neural networks, Robot. Auton. Syst. **51**, 81–99 (2005)

77.80 J.C. Houk, J.T. Buckingham, A.G. Barto: Models of the cerebellum and motor learning, Behav. Brain Sci. **19**(3), 368–383 (1996)

77.81 M.A. Arbib, C.C. Boylls, P. Dev: Neural models of spatial perception and the control of movement. In: *Cybernetics and Bionics*, ed. by W.D. Keidel, W. Handles, M. Spreng (Oldenbourg, Munich 1974) pp. 216–231

77.82 N. Schweighofer: Computational Models of the Cerebellum in the Adaptive Control of Movements, Ph.D. Thesis (University of Southern California, Los Angeles 1995)

77.83 N. Schweighofer, M.A. Arbib, M. Kawato: Role of the cerebellum in reaching quickly and accurately: I. A functional anatomical model of dynamics control, Eur. J. Neurosci. **10**, 86–94 (1998)

77.84 N. Schweighofer, J. Spoelstra, M.A. Arbib, M. Kawato: Role of the cerebellum in reaching quickly and accurately: II. A neural model of the intermediate cerebellum, Eur. J. Neurosci. **10**, 95–105 (1998)

77.85 M.A. Arbib, N. Schweighofer, W.T. Thach: Modeling the cerebellum: From adaptation to coordination. In: *Motor Control and Sensory-Motor Integration: Issues and Directions*, ed. by D.J. Glencross, J.P. Piek (North–Holland Elsevier Science, Amsterdam 1995) pp. 11–36

77.86 D. Wolpert, M. Kawato: Multiple paired forward and inverse models for motor control, Neural Netw. **11**, 1317–1329 (1998)

77.87 O. Donchin, J.T. Francis, R. Shadmehr: Quantifying generalization from trial–by–trial behavior of adaptive systems that learn with basis functions: Theory and experiments in human motor control, J. Neurosci. **23**(27), 9032–9045 (2003)

77.88 R. Osu, H. Gomi: Multijoint muscle regulation mechanisms examined by measured human arm stiffness and EMG signals, J. Neurophysiol. **81**(4), 1458–1468 (1999)

77.89 H. Höppner, J. McIntyre, S.P. der van: Task dependency of grip stiffness – A study of human grip force and grip stiffness dependency during two different tasks with same grip forces, PLOS ONE **8**(12), e80889 (2013)

77.90 P. van der Smagt, F. Groen, K. Schulten: Analysis and control of a rubbertuator robot arm, Biol. Cybern. **75**(4), 433–440 (1996)

77.91 G. Rizzolatti, G. Luppino, M. Matelli: The organization of the cortical motor system: New concepts, Electroencephalogr. Clin. Neurophysiol. **106**, 283–296 (1998)

77.92 G. Rizzolatti, L. Fogassi, V. Gallese: Neurophysiological mechanisms underlying the understanding and imitation of action, Nat. Rev. Neurosci. **2**, 661–670 (2001)

77.93 M.A. Umiltà, E. Kohler, V. Gallese, L. Fogassi, L. Fadiga, C. Keysers, G. Rizzolatti: I know what

you are doing: A neurophysiological study, Neuron **31**(1), 155–165 (2001)

77.94 P.F. Ferrari, S. Rozzi, L. Fogassi: Mirror neurons responding to observation of actions made with tools in monkey ventral premotor cortex, J. Cogn. Neurosci. **17**, 212–226 (2005)

77.95 L. Fogassi, P.F. Ferrari, B. Gesierich, S. Rozzi, F. Chersi, G. Rizzolatti: Parietal lobe: From action organization to intention understanding, Science **308**(4), 662–667 (2005)

77.96 M. Taira, S. Mine, A.P. Georgopoulos, A. Murata, H. Sakata: Parietal cortex neurons of the monkey related to the visual guidance of hand movement, Exp. Brain Res. **83**, 29–36 (1990)

77.97 A.H. Fagg, M.A. Arbib: Modeling parietal–premotor interactions in primate control of grasping, Neural Netw. **11**, 1277–1303 (1998)

77.98 G. Rizzolatti, G. Luppino: Grasping movements: Visuomotor transformations. In: *The Handbook of Brain Theory and Neural Networks*, 2nd edn., ed. by M.A. Arbib (Bradford Books/MIT Press, Cambridge 2003) pp. 501–504

77.99 S.H. Johnson-Frey: The neural bases of complex tool use in humans, Trends Cogn. Sci. **8**, 71–78 (2004)

77.100 G. Rizzolatti, L. Fadiga, V. Gallese, L. Fogassi: Premotor cortex and the recognition of motor actions, Cogn. Brain Res. **3**, 131–141 (1995)

77.101 E. Oztop, M.A. Arbib: Schema design and implementation of the grasp-related mirror neuron system, Biol. Cybern. **87**(2), 116–140 (2002)

77.102 D.I. Perrett, J.K. Hietanen, M.W. Oram, P.J. Benson: Organization and functions of cells in the macaque temporal cortex, Philos. Trans. R. Soc. B **335**, 23–50 (1992)

77.103 D.P. Carey, D.I. Perrett, M.W. Oram: Recognizing, understanding and producing action. In: *Handbook of Neuropsychology: Action and Cognition*, Vol. 11, ed. by M. Jeannerod, J. Grafman (Elsevier, Amsterdam 1997) pp. 111–130

77.104 B. Bonaiuto, E. Rosta, M.A. Arbib: Extending the mirror neuron system model, I. Audible actions and invisible grasps, Biol. Cybern. **96**(1), 9–38 (2006)

77.105 Y. Demiris, M.H. Johnson: Distributed, predictive perception of actions: A biologically inspired robotics architecture for imitation and learning, Connect. Sci. **15**(4), 231–243 (2003)

77.106 M.A. Arbib, A. Billard, M. Iacoboni, E. Oztop: Synthetic brain imaging: Grasping, mirror neurons and imitation, Neural Netw. **13**(8/9), 975–997 (2000)

77.107 S. Schaal, A.J. Ijspeert, A. Billard: Computational approaches to motor learning by imitation, Philos. Trans. R. Soc. Biol. Sci. **358**(1431), 537–547 (2003)

77.108 E. Oztop, M. Kawato, M.A. Arbib: Mirror neurons and imitation: A computationally guided review, Neural Netw. **19**(3), 254–271 (2006)

77.109 J. Tani, M. Ito, Y. Sugita: Self-organization of distributedly represented multiple behavior schemata in a mirror system: Reviews of robot

77.110 Y. Kuniyoshi, Y. Yorozu, M. Inaba, H. Inoue: From visuomotor self learning to visual imitation – a neural architecture for humanoid learning, IEEE Int. Conf. Robot. Autom. (ICRA), Vol. 3 (2003) pp. 3132–3139

77.111 D.M. Wolpert, K. Doya, M. Kawato: A unifying computational framework for motor control and social interaction, Philos. Trans. R. Soc. Biol. Sci. **358**(1431), 593–602 (2003)

77.112 G. Metta, G. Sandini, L. Natale, L. Craighero, L. Fadiga: Understanding mirror neurons: A biorobotic approach, Interact. Stud. **7**(2), 197–232 (2006) special issue on epigenetic robotics

77.113 D.M. Wolpert, Z. Ghahramani, R.J. Flanagan: Perspectives and problems in motor learning, Cogn. Sci. **5**(11), 487–494 (2001)

77.114 M.I. Jordan, D.E. Rumelhart: Forward models: Supervised learning with a distal teacher, Cogn. Sci. **16**(3), 307–354 (2006)

77.115 M. Kawato, K. Furukawa, R. Suzuki: A hierarchical neural network model for control and learning of voluntary movement, Biol. Cybern. **57**, 169–185 (1987)

77.116 M. Haruno, D.M. Wolpert, M. Kawato: MOSAIC model for sensorimotor learning and control, Neural Comput. **13**, 2201–2220 (2001)

77.117 E. Oztop, D.M. Wolpert, M. Kawato: Mental state inference using visual control parameters, Cogn. Brain Res. **22**, 129–151 (2005)

77.118 A.L. Woodward: Infant selectively encode the goal object of an actor's reach, Cognition **69**, 1–34 (1998)

77.119 M. Lopas, J. Santos-Victor: Visual learning by imitation with motor representations, IEEE Trans. Syst. Man Cybern. B **35**(3), 438–449 (2005)

77.120 A. Gijsberts, T. Tommasi, G. Metta, B. Caputo: Object recognition using visuo-affordance maps, IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS2010) Taipei, Taiwan (2010) pp. 1572–1578

77.121 N. Noceti, B. Caputo, C. Castellini, L. Baldassarre, A. Barla, L. Rosasco, F. Odone, G. Sandini: Towards a theoretical framework for learning multi-modal patterns for embodied agents, ICIAP-09, 15th, Int. Conf. Image Anal. Process. (2009)

77.122 P. Fitzpatrick, G. Metta: Grounding vision through experimental manipulation, Philos. Trans. R. Soc. Math. Phys. Eng. Sci. **361**(1811), 2165–2185 (2003)

77.123 S.M. Khansari Zadeh, A. Billard: Learning stable non-linear dynamical systems with Gaussian mixture models, IEEE Trans. Robotics **27**(5), 943–957 (2011)

77.124 A. Ude, A. Gams, T. Asfour, J. Morimoto: Task-specific generalization of discrete and periodic dynamic movement primitives, IEEE Trans. Robotics **26**(5), 800–815 (2010)

77.125 E.A. Theodorou, J. Buchli, S. Schaal: A generalized path integral control approach to reinforcement learning, J. Mach. Learn. Res. **11**, 3137–3181 (2010)

77.126 C.L. Nehaniv: Nine billion correspondence problems. In: *Imitation and Social Learning in Robots*,

Humans, and Animals: Behavioural, Social and Communicative Dimensions, ed. by C.L. Nehaniv, K. Dautenhahn (Cambridge Univ. Press, Cambridge 2006)

77.127   C.L. Nehaniv, K. Dautenhahn: Mapping between dissimilar bodies: Affordances and the algebraic foundations of imitation, Eur. Workshop Learn. Robots (EWRL-7) Edinburgh (1998)

77.128   Y. Demiris, G. Hayes: Imitation as a dual-route process featuring predictive and learning components: A biologically-plausible computational model. In: *Imitation in Animals and Artifacts*, ed. by K. Dautenhahn, C. Nehaniv (MIT Press, Cambridge 2002)

77.129   M. Ito, K. Noda, Y. Hoshino, J. Tani: Dynamic and interactive generation of object handling behaviors by a small humanoid robot using a dynamic neural network model, Neural Netw **19**(3), 323–337 (2006)

77.130   E.L. Sauser, A. Billard: Parallel and distributed neural models of the ideomotor principle: An investigation of imitative cortical pathways, Neural Netw. **19**, 285–298 (2006)

77.131   A.M. Liberman, F.S. Cooper, D.P. Shankweiler, M. Studdert-Kennedy: Perception of the speech code, Psychol. Rev. **74**, 431–461 (1967)

77.132   A.M. Liberman, I.G. Mattingly: The motor theory of speech perception revised, Cognition **21**, 1–36 (1985)

77.133   G. Rizzolatti, M.A. Arbib: Language within our grasp, Trends Neurosci. **21**(5), 188–194 (1998)

77.134   L. Fadiga, L. Fogassi, G. Pavesi, G. Rizzolatti: Motor facilitation during action observation: A magnetic stimulation study, J. Neurophys. **73**(6), 2608–2611 (1995)

77.135   M. Gentilucci, L. Fogassi, G. Luppino, M. Matelli, R. Camarda, G. Rizzolatti: Functional organization of inferior area 6 in the macaque monkey. I. Somatotopy and the control of proximal movements, Exp. Brain Res. **71**(3), 475–490 (1988)

77.136   F. Pulvermüller, L. Fadiga: Active perception: sensorimotor circuits as a cortical basis for language, Nat. Rev. Neurosci. **11**(5), 351–360 (2010)

77.137   L. Fadiga, L. Craighero, G. Buccino, G. Rizzolatti: Speech listening specifically modulates the excitability of tongue muscles: A TMS study. Eur. J, Neurosci. **15**(2), 399–402 (2002)

77.138   J.-L. Schwartz, A. Basirat, L. Ménard, M. Sato: The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception, J. Neurolinguist **25**(5), 336–354 (2012)

77.139   C. Moulin-Frier, M.A. Arbib: Recognizing speech in a novel accent: The motor theory of speech perception reframed, Biol Cybern. **107**(4), 421–447 (2013)

77.140   C. Castellini, L. Badino, G. Metta, G. Sandini, M. Tavella, M. Grimaldi, L. Fadiga: The use of phonetic motor invariants can improve automatic phoneme discrimination, PLoS ONE **6**(9), e24055 (2011)

77.141   A. D'Ausilio, F. Pulvermüller, P. Salmas, I. Bufalari, C. Begliomini, L. Fadiga: The motor somatotopy of speech perception, Curr. Biol. **19**(5), 381–385 (2009)

77.142   M. Grimaldi, F.B. Gili, F. Sigona, M. Tavella, P. Fitzpatrick, L. Craighero, L. Fadiga, G. Sandini, G. Metta: New technologies for simultaneous acquisition of speech articulatory data: Ultrasound, 3-D articulograph and electroglottograph, Poster, LangTech Conference, Roma (2008)

77.143   G.E. Hinton, V. Nair: Inferring motor programs from images of handwritten digits, Adv. Neural Inform. Proces. Syst., Vol. 18 (2004) pp. 515–522

77.144   L. Badino, C. Canevari, L. Fadiga, G. Metta: Deep-level acoustic-to-articulatory mapping for DBN-HMM based phone recognition, Spoken Language Technology Workshop (SLT) (2012) pp. 370–375

77.145   M.A. Arbib (Ed.): *From Action to Language Via the Mirror System* (Cambridge Univ. Press, Cambridge 2006)

77.146   C. Bell, P. Cordo, S. Harnad: Controversies in neuroscience IV: Motor learning and plasticity in the cerebellum, Behav. Brain Sci. **19**(3), v–vi (1996)

77.147   P. van der Smagt, D. Bullock: Applied intelligence, Scalable Appl. Neural Netw. Robotics **17**(1), 7–10 (2002)

77.148   V. Gallese, L. Fadiga, L. Fogassi, G. Rizzolatti: Action recognition in the premotor cortex, Brain **119**, 593–609 (1996)

77.149   L. Fadiga, L. Craighero, G. Buccino, G. Rizzolatti: Speech listening specifically modulates the excitability of tongue muscles: A TMS study, Eur. J. Neurosci. **15**(2), 399–402 (2002)