

# Sensory-Motor Primitives as a Basis for Imitation: Linking Perception to Action and Biology to Robotics

Maja J Matarić  
Computer Science Department and Neuroscience Program  
University of Southern California  
941 West 37th Place, SAL 228, MC 0781  
Los Angeles, CA 90089-0781  
mataric@cs.usc.edu

## X.1 Introduction

Imitation is a powerful means of skill acquisition, social interaction, and cultural transfer. Although the popular notion of "monkeying" and "parroting" makes the phenomenon appear ubiquitous, its more formal definition endows "true imitation" capability to very few species (Tomasello, Kruger & Rather 1993, Byrne & Russon 1998). The difficulty in teasing true imitation apart from related phenomena, including stimulus enhancement, social facilitation, emulation, and priming, indicates that the underlying mechanism for imitation is likely complex and related to the other types of social learning.

Given the power of imitation for individual and cultural learning, we consider it to be a rather broad capability that is based on a spectrum of behavior and learning of increasing complexity, only the most sophisticated of which correspond to the definition of "true imitation." In our work, we focus on imitation as the capability of acquiring new skills by observation, but allow for those skills to be combinations of the imitator's existing behavioral repertoire. More specifically, we focus on motor behaviors, and on their underlying mechanisms. In this chapter, we present a model of imitation that is based on a collection of direct sensory-motor mappings we call primitives. These primitives form a strong and specialized link between the visual and motor systems, and remain the active foundation for the more complex forms of imitation and other forms of social learning. A small set of such primitives may be innate (see Section X.5), and serves as a substrate for recognizing, classifying, and learning a broad repertoire of movement skills.

Our model is inspired and constrained by the following lines of evidence from psychophysics and neuroscience:

- 1) the existence of a selective attention mechanism for extracting movement information from the visual stream; this is based on data from attentional studies as well as our own eye-tracking work, described in Section X.4.1;
- 2) the existence of mirror neurons for sensory-motor integration/transformation; mirror neurons are a recently discovered mechanism for direct sensory-motor coordination, discussed in Section X.3;
- 3) the existence of motor primitives for structuring movement; motor primitives are a parsimonious means of encoding a basic movement vocabulary which can be composited into a broad and general movement repertoire, through the use of sequencing and superposition, as discussed in Section X.5;
- 4) the existence of a classification-based learning system, which uses the motor primitives to acquire new skills through composition (i.e., sequencing and superposition), and creation of new

representations, as described in Section X.7.3.

We propose an imitation model based on sensory-motor primitives, an evolutionarily old mechanism for integrating perception and action, which includes the function of mirror neurons, and results in a general movement vocabulary. The motivation and specific mechanisms of each of the parts of the model will be described in detail, and related to the relevant work in both natural and synthetic (e.g., robotic) imitation.

The rest of this chapter is organized as follows. We begin in Section X.2 with the motivation for this work, which bridges the fields of cognitive science and neuroscience with robotics. Next, in Section X.3, we survey the evidence and inspiration from neuroscience, cognitive science, and developmental psychology, and discuss mirror neurons, a key part of our model. In Section X.4, we review work on movement perception. We focus on eye-tracking work, including our own results, which are the basis for the attentional mechanism of the model. In Section X.5 we discuss motor control in biology (from a relevant neuroscience perspective) and robotics, focusing on motor primitives, another crucial component of our model. Section X.6 pulls together the described components of the model and also describes our experimental methods and test-beds. Section X.7 gives an overview of other robotic studies of imitation. Section X.8 summarizes the chapter and discusses continuing work.

## X.2 Motivation

Before describing *how* we model imitation, we first address our dual motivation for *why* we pursue it. We aim to elucidate a complex natural learning mechanism and use it to enable faster programming and adaptation in artificial systems, primarily robots. Imitation has clear relevance to both areas as it involves the integration of multiple cognitive/computational systems, namely perception, memory, action, and learning. This makes the endeavor challenging both in terms of analysis and synthesis, but its outcome is relevant to both biology and robotics.

Historically, the majority of work on biological imitation has focused on providing a precise classification of the phenomenon into distinct types and stages (Piaget 1962, Davis 1973, Tomasello et al. 1993) and addressing its role in social behavior (Bandura & Walters 1963, Bandura 1977, Meltzoff & Moore 1994). Recently, the evolution of imitation, and its relation to other forms of learning are being explored (Moore 1996, Moore 1992). Our work bears on the evolutionary history of imitation in that it outlines a potentially phylogenetically old mechanism as the core of the phenomenon.

Imitation is not only a powerful learning mechanism, but it is also thought to be a keystone in the evolution of communication and language (Arbib & Rizzolatti 1996, Rizzolatti, Gadicca, Gallese & Fogassi 1996, Jeannerod, Arbib, Rizzolatti & Sakata 1995, Donald 1993). Thus, imitation potentially vertically integrates cognitive systems from the lowest-level of perception and motor control to the highest levels of cognition. The notion of vertical integration is a foundational principle of *behavior-based robotics*, a new branch that aims at developing autonomous, adaptive, embodied systems that exist in the physical world and cope with the demands in the "here and now" (Brooks 1986, Brooks 1991, Matarić 1997). Unlike traditional robotics, where thinking (computation) takes more time than action, behavior-based robotics focuses on fully integrated systems which act in real time in their environment.

Imitation presents a very powerful mechanism for automated programming and control of robots. In the past, the idea of "teaching by showing" or "learning by demonstration" was limited to reproducing the goal or output state of the activity, but not the complete process that achieved it, or the underlying mechanism. In contrast, current efforts focus on modeling the imitation

mechanism itself, in order to develop a general tool for improved robot control.

In addition to understanding the body and the mind, a practical model of imitation has a variety of pragmatic real-world applications. These include more natural programming and instructing of robots, as well as improved human skill learning. We are particularly interested in those that aid skill (re)learning in a varied student population, ranging from children and athletes, to the handicapped and those recovering from injury.

### **X.3 Inspiration**

Evidence from neuroscience, cognitive science, and developmental psychology supports a strong link between the perceptual and motor systems involved in imitation. Decety (1996) and Jeannerod & Decety (1995) demonstrated that a shared neural substrate is used both in imagined and executed movements. Subjects involved in mental simulation (imagination and visualization) of motor tasks experienced activation of motor pathways similar to that during execution of movement. Consistent data were gathered with multiple modalities, including mental chronometry, autonomic responses, and cerebral blood flow experiments.

Similarly, increased activity in the motor cortex was measured during observation of movement, both in humans watching another human model (Jeannerod et al. 1995, Fadiga, Fogassi, Pavesi & Rizzolatti 1995) and in monkeys watching a human model (di Pellegrino, Fadiga, Fogassi & Rizzolatti 1992). These results also directly link perception of movement to its planning and execution. This link is supplied by so called “mirror neurons” found in area F5 of the monkey pre-motor cortex (Rizzolatti et al. 1996), which fire in response to observing specific movements as well as during execution of the same movement. Similar facility was found in the Broca’s area of the human brain. Since their discovery, mirror neurons have become a subject of intense study due to their intriguing integration role between perception and action, and thus they are a natural candidate for neural theories of imitation. However, it is important to note that mirror neurons were found in monkeys, which are not considered capable of “true imitation”; the presence of mirror neurons in other species has not yet been tested extensively. We postulate that the function mirror neurons perform, the connection between observed and performed behavior, is the substrate for imitative capability (and possibly for communicative capabilities, as proposed by Arbib & Rizzolatti (1996)), since simpler forms of mimicry are manifested by those animals.

The neuroscientific evidence discussed so far has addressed the connection between sensory and motor systems. Psychophysical experiments addressing a cognitive aspect of this connection provide complementary evidence. Vogt (1995) tested human subjects for the existence of an intermediate processing/coding level between perception and action during an imitation task. The subjects were shown sine and cosine drawings and asked to draw them with their eyes closed. They showed no improvement after immediate rehearsal, and no effect from distractions. This result supports a lack of an intermediate coding level and the presence of a generative perception module, i.e., one that perceives in such a way as to prepare for action. Similarly, Pelisson, Goodale & Prablanc (1978) showed that adjustments of hand pointing to visual targets do not require visual information about the moving arm.

Further evidence supporting sensory-motor integration (also viewed as generative perception), can be found in developmental psychology studies (Bertenthal 1996), lesion studies (Goodale, Jakobson, Milner, Perrett, Benson & Hietanen 1994), and studies of athletic performance, in skills such as squash and ball catching (Abernethy 1990, Elkins 1996, Savelsbergh, Whiting, Pijpers & van Santvoord 1993, Savelsvertgh & Bootsma 1994).

Several decades of developmental psychology work in infant imitation, conducted by Meltzoff

and Moore, also actively support sensory-motor integration as the basis of imitation. Meltzoff & Moore (1988) and Meltzoff & Moore (1983) argue that imitation is a fundamental human capability, found in newborns and based on an innate link between the perceptual and motor systems (Meltzoff & Moore 1994). This link enables young children to imitate facial expressions and hand movements without visual feedback. The authors hypothesize innate, integrated, supra-modal representations of human movements and postures as a part of the imitation mechanism (Meltzoff & Moore 1995, Meltzoff & Moore 1997).

The above and other related biological evidence serves as powerful inspiration for our philosophy and approach to modeling imitation. It motivates the four key principles of our work, each now addressed in turn.

## **X.4 Movement Perception and Understanding**

We now turn to movement perception, and give an overview of the selected biological evidence and the robotics approaches relevant to the imitation problem.

### **X.4.1 Biological Evidence**

Perrett and co-authors performed a series of studies of movement perception in the macaque monkey. These resulted in a consistent hypothesis of the existence of specialized neural detectors (and predictors) for specific postures and movements and goal-oriented behaviors. Perrett, Smith, Mistlin, Chitty, Head, Potter, Broennimann, Milner & Jeeves (1985) describe data from the macaque, showing neurons that selectively respond to specific movement types and stimulus forms, i.e., for specific body and body part movements. Perrett, Harries, Bevan, Thomas, Benson, Mistlin, Chitty, Hietanen & Ortega (1989*a*) demonstrate similar, view-independent results, implying the potential for viewer-centered and goal-centered descriptions of the movement. Perrett, Harries, Mistlin & Chitty (1989*b*) describe a population of neurons involved in the recognition of specific actions of others, and Perrett, Harries, Mistlin, Hietanen, Benson, Bevan, Thomas, Oram, Ortega & Brierley (1990) show that expectation of movement indirectly correlates with the amount of resulting neural firing.

Taken together, this evidence supports the presence of specialized, integrated movement detection system. As discussed below (Section X.5), this evidence fits well with the motor primitives component of our model.

### **X.4.2 Robotic Approaches to Movement Understanding**

Computational theories and implementations of movement perception have dealt with classifying movement largely by employing time-series analysis tools. The most used tool are Hidden Markov Models (HMMs) (Rabiner & Juang 1986), originally used to automatically segment and recognize speech (Rabiner 1989), but more recently generalized to human and robot movement analysis. For example, Pook & Ballard (1993) use it to analyze a video of a robot flipping a plastic fried egg. More recently, Brand, Oliver & Pentland (1997) describe a coupled Hidden Markov Model for recognizing and analyzing time-extended behaviors such as object assembly. Lee & Xu (1996) used a similar HMM-based approach to recognize several hand-shaped letters. Yang, Xu & Chen (1997) applied the HMM representation to model human skill learning by observation and skill transfer to articulated robot arms in two task domains: 1) recognition and copying of written digits and 2) parts replacement by tele-operation.

More recently, the field of human-computer interfaces and its various relatives have addressed movement perception and greatly broadened the spectrum of approaches being used. People tracking and gesture recognition from video are two major applications of interest, and most closely related to the work described here. Movement perception is a very challenging problem, and the tools that have been explored to deal with it are varied. They include color, models, and features, often used in combination, and usually at a high computational cost. A detailed survey of movement perception is beyond the scope of this chapter; for more information see Essa (1999).

#### **X.4.1 Eye-Tracking Studies and Results**

Johansson (1977) demonstrated the human capability to recognize biological motion from a small number of structured visual cues such as light dots in key locations on the limbs. This well-known and subsequently often replicated experiment points toward a mechanism for body-motion detection and understanding. To gain insight into what humans attend to while observing movement, eye-trackers can be used to measure overt fixation locations. Historically, eye-tracking studies have largely been performed on static images, applied to tasks such as skill learning from picture-text combinations (Daugis, Bliscke & Olivier 1978) and object boundary tracking (Kudo, Uomori, Yamada, Oinishi & Sugie 1992). They have also been used to study working memory representations and limitations. (Ballard, Hayhoe, Pook & Rao 1997) used a head-mounted eye-tracker to record saccades in an imitation-style task that involved assembling a toy aeroplane from observation (Ballard et al. 1997). In all the above, the task and thus the output state was more important than the behavior that achieved it, so these were not studies of "true imitation."

Based on our interest in imitation, and the focus on visuo-motor integration, we conducted a set of eye-tracking experiments aimed at addressing the following questions:

1. Is there a difference between watching to imitate versus just watching?
2. When watching to imitate, what features are fixated on?

In a standard cognitive model of perception and motor control, one would expect the answer to the first question above to be positive. However, if we assume strong sensory-motor integration assumption, we would expect to see no difference in fixation patterns between the imitation and no-imitation conditions, since the unconscious process of early stages of movement preparation would be uniformly active in both. Our results provided a decisive negative answer to the first question, thus adding overt fixation behavior to other data that support the sensory-motor connection.

The second question was aimed at gaining insight into the underlying mechanisms for motor control. It studied the sparse information provided by fixation data and its use in subsequent movement imitation, which humans routinely perform with excellent proficiency (Tomasello et al. 1993, Kolb & Milner 1981, Davis 1973). In our experiment, 40 subjects had their fixations recorded while watching 25 3-second long videos of different types of unfamiliar finger, hand, and arm movements, shown on a computer screen. Before each set of videos, the subject were told whether they would subsequently imitate each video after viewing it. The computer screen was black between the stimuli and during the imitation phase. In the imitation condition, the subject extended the right arm and imitated without visual feedback.

Regardless of the type and size of presented stimulus, and regardless of whether they were only watching or would subsequently imitate what they saw, subjects looked most often and for the longest time periods at the hand; fixation frequency and duration were both maximized at the endpoint. The only difference between the subjects' behavior between the imitation and no-imitation conditions was in pupil dilation. In the imitation condition, dilation was significantly larger, while the fixation behavior remained the same in both conditions. Finally, in spite of focusing overt attention on the end-point, the subjects were effective at imitating the presented movements. For

a detailed description of the experiments, analysis, and results, see Matarić & Pomplun (1998).

The eye-tracking results support Meltsoff and Moore’s developmental work, Perrett’s work on movement perception, and are consistent with a growing body of neuroscience evidence for generative perception of movement, whether it be specific to imitation, or as a general aspect of visuo-motor processing.

More specifically, the results demonstrating consistent fixations at the end-point allow for several hypotheses that impact our imitation model:

1. that an end-effector-driven attentional mechanism may exist, greatly simplifying the overt attention problem;
2. that internal models (possibly in the form of primitives) are being used to fill in the motor control details which are not directly observed;
3. that those internal models also influence attention in the form of providing movement prediction or expectation;
4. and finally that an efficient transformation between the two representations (and their coordinate-frames, i.e., the extrinsic end-point position and the intrinsic joint angles of the arm) must exist in order to convert the observed end-point information into executable control for the arm.

Before addressing how these hypotheses are integrated into our model, we first turn to motor control and its organization and relation to perception.

## **X.5 Motor Control and Motor Primitives**

Understanding the modular organization of biological movement and implementing it in artificial systems are outstanding challenges in biology and robotics. The theory of motor primitives as the underlying representation of movement has served as an inspiration and model for work both in behavior-based control in general and in robot models of imitation in specific.

The existence of force-field motor primitives in the spine, which converge to single equilibrium points and produce high-level behaviors such as reaching and wiping, has been suggested (Giszter, Mussa-Ivaldi & Bizzi 1993, Mussa-Ivaldi & Giszter 1992). When the spine of a frog is stimulated with an electrode, a particular field is activated, the frog’s leg executes a behavior such as wiping, and comes to rest at a position that corresponds to the equilibrium point of that field. Furthermore, when two or more fields are stimulated at the same time, either a linear superposition of the fields is obtained, or one of the fields dominates (Mussa-Ivaldi, Giszter & Bizzi 1994). In either case, a meaningful movement results. Careful studies found only a small number of such distinct fields in the frog’s spine. Thus only a dozen primitives may be necessary and sufficient for coding the frog’s entire motor repertoire, through sequencing and superposition of supra-spinal inputs (Bizzi, Mussa-Ivaldi & Giszter 1991).

The above suggests an elegant and modular organizational principle for motor control, in which entire behaviors are coded with low-level force-fields. Through combination of individual fields, higher level behaviors can easily be synthesized. Abstracting away from the specific coding of the spinal fields, the examples from neurobiology provide the framework for a motor control system based on a small number of additive primitives (or basis behaviors) sufficient for a rich output movement repertoire. Our previous work (Matarić 1995, Matarić 1997), inspired by the same biological results, has successfully applied the idea of basis behaviors to control of mobile robots

by fitting it directly into the modular behavior-based control paradigm. Applications of schema theory (Arbib 1992) to behavior-based mobile robots (Arkin 1987) have employed a similar notion of composable behaviors, stemming from foundations in neuroscience (Arbib 1981, Arbib 1989).

The idea of using such primitives for articulator control has been recently studied in robotics. Williamson (1996) and Marjanović, Scassellati & Williamson (1996) developed a 6 DOF (degrees of freedom) robot arm controller. While in the biological and mobile robotics work primitives code for behaviors, theirs coded for four static postures. Three reaching and one resting posture were used and interpolation was applied to reach any goal end-point position.

Matarić, Zordan & Williamson (1999) used one of the experimental test-beds from the imitation work described here (see Section X.7), the dynamical humanoid simulator, to implement two versions of the force-field primitives to control the 7 DOF humanoid arms. One implementation, closely modeled on the frog work, used an intrinsic, joint-space representation. Another motor controller implementation used another biologically-inspired approach, impedance control (Hogan 1985), which operates in the extrinsic end-point (hand) coordinate frame. Both were demonstrated to provide an effective substrate for control of a complicated sequential motor task, but each had limitations in terms of ease of control for certain types of movements. Neither reference frame was uniformly more convenient and thus a combination solution was proposed.

The issue of coordinate frame choice is just one of the many involved in the primitives-based approach. While it is intuitively clear why an expressive and additive vocabulary of primitives is an appealing idea, the question remains open of *what* the primitives in such a vocabulary should be. The problem becomes even more complex when we combine it with perception. The choice of primitive is one of the major issues in our imitation model, addressed in detail in Section X.7.2. Before presenting our imitation model, we briefly review other robotics work on imitation.

## X.6 Imitation in Robotics

The earliest robotics work to address imitation was focused on assembly task-learning from observation. Typically, a series of visual images of a human performing a simple object moving/stacking tasks was recorded, segmented, interpreted, and then repeated by a robotic arm (Ikeuchi, Suehiro, Tanguy & Wheeler 1990, Ikeuchi, Kawade & Suehiro 1993, Kuniyoshi, Inaba & Inoue 1994, Holland, Sikka & McCarragher 1996, Kaiser 1997). This work was aimed at repeating the final outcome of the observed behavior, not at imitating the process that brought the result about. Thus, the goal was to achieve task-level imitation (Byrne & Russon 1998), the main focus being on extracting the task from the visual data, in the form of intermediate states and sub-goals. Once the sub-goals were extracted, a conventional trajectory planner was usually employed to reach each of them (for example: pick up block1, put it on top of block2). These efforts constitute a significant body of research in robotics, and contribute to video segmentation and understanding. However, they do not address imitation as defined in the biological literature.

Schaal (1997) applied imitation to “priming” a model-based reinforcement learning system in the task of pole balancing by a 7 DOF robot arm. After a brief (30 second) demonstration by a human, the system learns the task from a single trial. Demiris & Hayes (1997) used a robotic head equipped with a pair of cameras that observed and imitated a set of head movements of a human demonstrator (Demiris, Rougeaux, Hayes, Berthouze & Kuniyoshi 1997). The approach used a visual feature detector, which informed a built-in system that directly mapped a set of possible observed head movements to the robot’s own movements. The visual feature detectors were inspired by Perrett’s specialized neurons (Perrett et al. 1985, Perrett et al. 1989a) and the observed-performed mapping by Meltzoff’s innate visuo-motor map (Meltzoff 1993, Meltzoff &

Moore 1994). Using a mobile robot Hayes & Demiris (1994) employed a similar direct-mapping imitation mechanism to learn maze navigation, inspired by work described in Dautenhahn” (1995). Recent efforts, including our own (Matarić 1994), have been increasingly oriented toward analyzing the underlying mechanisms of imitation in natural systems and modeling those on artificial ones.

## X.7 Our Model and Implementations

Attempts to endow robots with imitation capabilities are quite recent, and those aimed at studying biological imitation through robotics are few. Our work is an attempt to combine information from these two disparate fields to create an interdisciplinary alliance to address the complex phenomenon.

As described above, the key biological inspirations for our approach to modeling come from evidence of sensory-motor integration in perception, from mirror neurons, and from motor primitives for control. We combine those elements into a model with the following key components:

- 1) a selective attentional mechanisms for extracting salient movement information from the visual stream by focusing on the end-points;
- 2) sensory-motor primitives as a means of representational integration and transformation, combining the roles of mirror neurons and motor primitives for structuring movement;
- 3) a classification-based learning mechanism that utilizes the primitives to acquire new skills through composition (sequencing and superposition) and creation of new representations.

These components constitute the model of imitation that we are experimentally validating on several different testbeds (see Section X.7.4). Each of the three main model components is described next.

### X.7.1 Tracking of Movement and Visual Attention

Deciding what to pay attention to for subsequent imitation it is a difficult problem. Based the literature on visual attention, and on our own eye-tracking results, we believe that the focus of attention is not driven directly by the imitation goal. Instead, we postulate an attentional mechanism that takes into consideration both intrinsic bottom-up features (such as velocity changes, and the position of the end-point) and extrinsic, top-down features mandated by the task (imitation or otherwise).

We first employ a motion tracking system capable of selecting a collection of features from the moving image, based on a constrained (un-occluded and unambiguous) initial position and kinematic model of the object/body being imitated. (Our work so far assumes that a human is being imitated, and the kinematic model used is that of a generic adult human. However, the system is general and can use any kinematic model.) This greatly simplifies establishing the initial match between the features in the visual image and the underlying body. The match enables tracking of the body over time, using a planar rectangle representation of the limbs. This allows for fast computation and updating of limb position, and for simple prediction of future position, which is in turn used to speed up recognition. (Input from the motor primitives could provide further improved predictive capability, but we are yet to properly explore it.) This approach provides movement tracking in a very computationally inexpensive fashion because it almost entirely rests on image processing, only employing a simple, stick-figure-based kinematic model for the initial match, but not any explicit dynamic models of the observed body to predict future positions (Bernardo, Goncalves & Perona 1995, Gavrila & Davis 1995).

The system must next decide which of the features provided by the movement tracking system should be attended to. This decision is made by the attention selection mechanism, based on a combination of task-driven and intrinsic biases. We postulate that intrinsic factors that draw

attention are changes in speed and direction of an object in the visual field. Additionally, any held objects and end-points of manipulators (i.e., hands, fingers) also intrinsically draw attention. The latter are interesting for a combination of reasons: they are typically (but not always) the fastest moving parts of the body, and are usually most directly involved with achieving the task at hand.

In our implementation, the attention selection process, i.e., the selection of features to attend to, consists of the following by three stages:

1) Segmentation of retinal space; 2) Evaluation of current interest of each feature; 3) Displacement of the center of attention.

In a preprocessing step, all of the features are transformed into an egocentric coordinate frame (Stein 1992, Flanders, Tillery & Soechting 1992, Flanders & Soechting 1995) suitable for subsequent higher-level processing of the input. The egocentric coordinate axes are found by either arbitrarily selecting an anchor point based on the extracted features or through motion tracking a visual landmark.

Our segmentation model is loosely based on the organization of the human retina. There, the fovea is the center of attention and provides the highest resolution of the original image. As the distance between a feature and the center of attention (the fovea) increases, the resolution of the feature's perceived position decreases. To model this effect, we use a log-polar grid (Hubel & Wiesel 1977) to perform retinal image segmentation. Each cell in the grid represents a uniform spatial region; if a feature is within a cell, its perceived position is the cell's center.

The evaluation of the current interest of a feature is based on three factors: retinal movement, captivation, and task-level bias. Retinal movement is the amount a feature has recently moved in the retinal image. Captivation of a feature is based on its recent past interest level and the overall level of "boredom". Boredom is inversely proportional to the number of newly-active retinal cells, i.e., to the novelty of the image. A feature becomes boring when it is given attention and produces little or no newly occupied cells in the retina. Intuitively, changes in movement reduce boredom and draw attention; a moving feature is interesting for a while, but loses interest if its movement does not change over time. Finally, task-level bias is any preference for specific features introduced by the task itself (e.g., the feet for learning dance steps), encoded as an additional weight factor.

To compute the total interest level of a feature, the inputs from the three factors are normalized and weighted as follows: retinal movement is the basic driver of attention, captivation can supersede it provided sufficient activation, and task-level bias can override both also when provided sufficient activation.

The three factors above represent dimensions in feature-interest space. Each feature is represented as a point in that space, with its magnitude corresponding to its interest level. Displacement of the center of attention is based on the computed interest levels of the features and their distance, in interest space, from the current center of attention. A distance threshold is used to eliminate uninteresting features. If none are left, attention de-focuses on either the entire performer or a subset of recently interesting features. Otherwise, the remaining feature with the most interest captures the attention and inhibits the others.

In summary, the tracking system provides features and the attentional system selects the most interesting one(s) to attend to. These selected features serve as the basis for movement classification into primitives.

### **X.7.2 Sensory-Motor Integration Through Primitives**

We have proposed primitives as the unifying mechanism between the perceptual and motor systems. We have also presented biological evidence for motor primitives, which have been successfully applied to robot control as well. However, we have not yet discussed how perceptual information connects

to such motor primitives.

This is one of the many challenges of the proposed model. On one hand, the primitives are driven by the constraints of the motor system (i.e., its kinematics and dynamics) and the task (frequently executed movements). On the other hand, they are influenced by the structure and inputs into the visual system. The motor primitives system serves as an effective substrate for structuring movement, regardless of the organization of perception. However, if we postulate a primitives-based integration of perception and motor control, then the perceptual inputs must also be similarly structured and constrained. In this section we give an example of such a primitives-based perceptual system, which is mapped to a matching motor control system.

Before proceeding, we address the issue of perceptual primitives. Unlike the motor system, whose output is kinematically and dynamically constrained, the input into the perceptual system is much less limited. In fact, it is impossible to classify all possible inputs into a useful set of task-independent categories. This is where imitation plays a crucial role: we postulate that the mirror neuron system selects relevant classes of observed *biological movement* and maps those directly to the motor primitives. Thus, not all visual input is so classified; the system is evolved for internal and external mimicry and imitation of observed movements of conspecifics and other relevant animals. Consequently, imitation across species is possible.

The representation of the primitives, including their frame of reference, is their most fundamental property, and has a critical effect on their subsequent usefulness. A key role of primitives is to integrate the perceptual representation, in retinotopic space, with motor representation, in intrinsic joint space. We have experimented with motor primitives represented in both intrinsic, joint space of the imitator, and in extrinsic (Cartesian) observer-based space (Matarić et al. 1999). As expected, different representations are more or less appropriate depending on the task specification and type of movement to be imitated. For example, tasks involving body postures are best expressed in intrinsic coordinates, while those involving end-point movement are best expressed in extrinsic space. This is intuitive, given the complexity of the human body and the limited resources of the attentional mechanism.

Using our eye-tracking results we postulate that most attention is focused at the end-point(s) of the model being imitated, while the posture is coarsely observed and interpolated based on the imitator’s own internal models of movement. When postures or body movements are to be imitated, task-based bias can draw attention to specific joints and other relevant body parts. This, however, makes for a harder imitation task, and requires more rehearsal and training. Not surprisingly, trained dancers and athletes are much better equipped for this type of imitation, whether it be due to their better attentional mechanisms or more extensive movement repertoires (Matarić & Pomplun 1998). From an evolutionary perspective, imitation of whole-body posture and movement appears to be less useful than imitation of end-effector behaviors, such as the use of tools, which may have more selective advantage. This reasoning leads us to postulate an evolutionarily old mechanism, such as mirror neurons, as a basis for sensory-motor integration, evolved or adapted for learning by observation and demonstration, culminating in the complex capability of imitation.

In our current model, sensory-motor primitives encode movements invariant to exact Cartesian position, rate of motion, size, and perspective. The primitives represent a basis set that is mapped to a set of parametric motor controllers that can be used to generate new, more complex motions. In this sense, all complex motions and skills are learned by sequencing and combining (superimposing) primitives, which were themselves either learned or innate. The origin of primitives is another challenge: while the term implies innateness, the need to accommodate a growing movement repertoire requires them to be either fully general or adaptive. We postulate a small set of innate primitives and a composition mechanism (based on sequencing and superposition) that allows for learning new behaviors which are themselves used as composable elements.

One may question how generic primitives can contain enough information for high-fidelity imitation. However, the notion of representing a complex function as a linear combination of much simpler basis functions is well established in mathematical physics (Levine 1983). The same motivation is used here. High-fidelity imitation requiring specific quantitative measurements of the observed movement (e.g., Cartesian position and configuration of the kinematic system), is still possible, since all of that information is available to the vision system, and can be reconstructed on-the-fly, but need not be stored permanently. Thus, when imitating a movement, the primitives determine its higher-level description, and any specific temporarily held metric information is used to parametrize that description into an executable movement, and fed directly to the controller.

The generic encoding of the primitives, which can be parametrized by more specific metric data at execution-time, greatly reduces the necessary parameter space of the primitives representation. At the same time, it allows the attentional mechanism to acquire any additional information needed for imitating a given demonstration and use it directly at the control level, without the need for permanent storage. From a robotics perspective, the notion of endowing a system with a collection of behavioral primitives and thus avoiding on-line run-time trajectory planning computation, is very appealing (Mataric et al. 1999).

Much of synthetic imitation work assumes a direct, one-to-one mapping between the perceived movements and a robot's own behavior repertoire. While this may correspond to a very small set of innate behaviors (Meltzoff 1993), it is unlikely to scale to full complex movement repertoires. In contrast, our work utilizes a classification mechanism to find the best fit between the observed behavior and combinations of primitives. *What are the right primitives?* To address this difficult question, we are experimenting with two types of systems: 1) those that are given a set of "innate" primitives, and 2) those that learn the primitives themselves. Our previous robotics work has addressed the process of designing primitives, or basis behaviors, for a specific system and set of tasks (Mataric 1995). As mentioned above, we have explored various representations, including convergent vector fields in both joint and Cartesian space, impedance control, interpolated joint-space control (Mataric et al. 1999), and central pattern generators (CPGs). Recently, we have begun exploring the other approach, namely learning the primitives themselves, described next.

### X.7.3 Classification and Learning

The very notion of a basis set of primitives implies that those are innate, and not adaptable. While that is a likely interpretation of spinal motor primitives (Bizzi et al. 1991), it is not as likely for more general sensory-motor mappings, such as the function of mirror neurons. Therefore, for the purposes of imitation in artificial systems, we consider the possibility that the primitives themselves can be learned, or at least adapted based on training.

It is intuitive that the nature of the primitives determines what types of movements will be most easily classified. We are interested in an general basis set, which will serve as a vocabulary to enable the imitation and generation of a broad repertoire of movements. In the absence of an obvious way to represent general movement parsimoniously, we choose a data-driven approach: the primitives are learned from a set of movement data. Since these data are training examples, their structure, number/size, and variability greatly influence what primitives will be extracted and learned. This is an unavoidable property of any learning systems, biological or synthetic. In our approach, we present a wide variety of simple movements to the learning system, and allow it to automatically extract the primitives.

The movements presented to the learning system, i.e., the training data, are presented in the same form as any input into the vision system, and are processed into a collection of features over time. Depending on the desired dimensionality and thus complexity of the learning problem,

different numbers of features can be used for learning.

In one of our approaches, we are using only end-point movement over time as the source of training data. Besides the already discussed evidence from eye-tracking and other studies, robotics also provides justification for this approach. For any given kinematic structure (i.e., robot arm), the most critical goal is typically to achieve the desired end-point position while satisfying a few constraints for the rest of the body (e.g., no collisions, arm configuration, etc.). Thus, it is pragmatic as well as computationally efficient to endow a robot with a set of behavioral primitives for achieving a set of parameterized end-point movements, and apply any additional execution-time constraints to those primitives. End-point position over a fixed temporal horizon (time-slice) is used as input to the learning system. Other information about the movement is ignored, but can be used for subsequent imitation (see Section X.7.2). The end-point position during a time-slice is encoded as a matrix of 2-dimensional movement vectors between two successive frames within the time horizon. The movements during a time-slice are averaged in the temporal dimension, to make them time-invariant, and any spatial outliers are removed. This process is applied to different temporal horizons, so that multiple time-scales may be captured.

In addition to exploring the effectiveness of end-point-based primitives, we are also applying clustering and classification techniques to higher-dimensional representations of the body. Specifically, we use a 9-dimensional feature vector, consisting of a head, torso, waist, and two shoulders, elbows, and hands, as input to the classification/learning system. We are currently exploring several different classification and learning algorithms in order to properly address the key aspects of this learning problem, especially the need for temporal and spatial invariance in the presence of strong temporal and spatial structure in the data. Specifically, the temporal sequence is a defining property of any movement, but its specific duration may not be. Additionally, the spatial structure of the kinematic model is also a fundamental property of the body, but the spatial location of the movement, unlike its overall pattern, is not.

Once the time- and space-invariant vector representations are computed, they are classified into clusters using vector quantization (Martinetz & Schulten 1991). Once such an invariant representation is achieved, any number of different clustering techniques can be applied. The resulting clusters represent perceptual primitives. Once learned, these serve as the vocabulary for classification of any observed movement. Once a movement is classified, the resulting composition of primitives is mapped onto a collection of motor control primitives capable of generating the observed movement.

It is important to note that learned primitives, while an effective basis for superposition, may not necessarily appear meaningful on their own. Depending on the amount of bias presented in the training data and on the learning mechanism, we can enforce that the learned primitives are themselves meaningful, if sequencing is expected to be their primary use.

We are also working on learning new skills as temporal sequences of movements. One of our approaches involves using a recurrent neural network architecture, DRAMA (Billard & Hayes 1999), with general abilities for learning complex time series.

Besides learning useful new sequences of primitives, some superpositions can be recognized as a good candidates for new primitives. For example, the measure of fit between the observed movement and the resulting primitives classification can be used for further learning. If the fit is poor, i.e., if the movement is distributed over most of the primitives, it may be used as an indication for the need for a new primitive. Additionally, frequent superpositions also form a natural basis for learning new composable elements.

The approach we have described maps a set of time- and space-invariant visual primitives onto a set of parametrized motor primitives or controllers. This arrangement allows for more flexibility for both systems, but involves a mapping between the primitives themselves. Our continued work is aimed at exploring various representations for the primitives and their effect on imitation.

Our model of imitation is being validated experimentally as it is developed. The experimental test-beds are described next.

#### X.7.4 Experimental Test-Beds

We are using four distinct experimental test-beds to validate various implementations of our imitation model.

*Physics-based humanoid simulation:* Adonis, a physics-based humanoid simulation<sup>1</sup>, is actuated from the waist up, and has 20 degrees of freedom (DOF): 3 in the neck, 3 in the waist, and 7 in each arm (3 in the shoulder and wrist, and one in the elbow). Adonis uses external sensing (such as vision input), and internal joint angle state sensors. This test-bed was used for our past motor control work and continues to be the closest test-bed to human body control.

*Humanoid avatars:* Two COSIMIR<sup>2</sup> humanoid avatars are fully actuated and consist of 37 body DOF and 14 hand DOF in actuated fingers. Various models of the dynamics, muscles, and control can be applied. All sensing is external.

*Robot dog:* Laika, an Aibo<sup>3</sup> dog robot, has a total of 18 degrees of freedom: 2 in the neck and 4 in each leg. It also has passively compliant paws. Its sensors include a color CCD camera, stereo microphones, joint encoders in the legs, and touch sensors on the paws and the head. The main reason for using Laika is to study imitation between different morphologies. Laika will be used to test our imitation model on imitating a human. Additionally, Laika's movements will be used as input to the other experimental test-beds.

*Wheeled mobile robots:* Increasing numbers (currently 6) of Pioneers<sup>4</sup>, mobile robots equipped with pan and tilt color cameras, front and rear ultrasound sensors, and differential steered bases, are used for group robotics research in our laboratory. Although these bases are too different from biological ones to be used for studying the direct mechanisms of true imitation, they are very convenient for addressing task-level imitation, which is a challenging problem in its own right given the limitations of robot sensors and effectors.

The purpose behind using such a broad variety of test-beds is to validate the generality of our model, as well as put it practical use on various complex agents and robots with potential real-world programming and control applications.

### X.8 Summary and Continuing Work

We have presented a model of imitation based on evolutionarily old mechanisms of mirror neurons and motor primitives. Combining evidence from neuroscience, cognitive science, and developmental psychology, we have postulated that that the model uses selective visual attention driven by end-point movement, a set of sensory-motor primitives representing a basis movement vocabulary, and a classification and learning system for creating novel sequences and superpositions of the primitives to continually expand the imitator's movement repertoire. We continue to develop and validate this model on a variety of robotic test-beds, and hope to provide both novel insight into biological imitation and novel techniques for programming and control of robots.

---

<sup>1</sup>Developed in the Animation Lab at Georgia Institute of Technology, using S/D Fast.

<sup>2</sup>Developed at the University of Dortmund, described in Romann (1999)

<sup>3</sup>Developed and distributed by Sony.

<sup>4</sup>Distributed by ActivMedia.

## Acknowledgements

The work described here is supported by the National Science Foundation Career Award IRI-9624237. The vision system was developed and implemented by Stefan Weber, who continues the work on learning perceptual primitives. The work on attentional selection was developed and implemented by Odest Chadwicke Jenkins. Work on sequence and symbolic learning is being worked on by Aude Billard. The author thanks all of the above and especially Richard Roberts for invaluable comments on earlier drafts of this chapter.

## X.9 Bibliography

### References

- Abernethy, B. (1990), ‘Expertise, visual search, and information pick-up in squash’, *Perception* **19**, 63–78.
- Arbib, M. (1981), Perceptual Structures and Distributed Motor Control, in V. B. Brooks, ed., ‘Handbook of Physiology: Motor Control’, The MIT Press, pp. 809–813.
- Arbib, M. (1992), Schema Theory, in S. Shapiro, ed., ‘The Encyclopedia of Artificial Intelligence, 2nd Edition’, Wiley-Interscience, pp. 1427–1443.
- Arbib, M. & Rizzolatti, G. (1996), ‘Neural Expectations: A Possible Evolutionary Path From Manual Skills to Language’, *Communication and Cognition* **29**(2–4), 393–424.
- Arbib, M. A. (1989), Visuomotor Coordination: Neural Models and Perceptual Robotics, in ‘Visuomotor Coordination: Amphibians, Comparisons, Models and Robots’, Plenum Press, pp. 121–171.
- Arkin, R. C. (1987), Motor Schema Based Navigation for a Mobile Robot: An Approach to Programming by Behavior, in ‘IEEE International Conference on Robotics and Automation’, Raleigh, NC, pp. 264–271.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K. & Rao, R. P. N. (1997), ‘Deictic Codes for Embodiment of Cognition’, *Behavior and Brain Science*.
- Bandura, A. (1977), *Social Learning Theory*, Prentice-Hall, Inc., Englewood Cliffs, N.J.
- Bandura, A. & Walters, R. H. (1963), *Social Learning and Personality Development*, Holt, Rinehart and Winston, Inc, New York.
- Bernardo, E. D., Goncalves, L. & Perona, P. (1995), Monocular Tracking of the Human Arm in 3D: Real-Time Implementation and Experiments, Technical report, California Institute of Technology and Universita di Padova, Italy.
- Bertenthal, B. I. (1996), ‘Origins and early development of perception, action, and representation’, *Annual Review of Psychology* **47**, 431–459.
- Billard, A. & Hayes, G. (1999), ‘DRAMA, a connectionist architecture for control and learning in autonomous robots’, *Adaptive Behaviour Journal* **7**(1), 35–64.

- Bizzzi, E., Mussa-Ivaldi, F. A. & Giszter, S. (1991), ‘Computations Underlying the Execution of Movement: A Biological Perspective’, *Science* **253**, 287–291.
- Brand, M., Oliver, N. & Pentland, A. (1997), Coupled hidden Markov models for complex action recognition, in ‘Proceedings, CVPR’, IEEE Press, pp. 994–999.
- Brooks, R. A. (1986), ‘A Robust Layered Control System for a Mobile Robot’, *IEEE Journal of Robotics and Automation* **RA-2**, 14–23.
- Brooks, R. A. (1991), Intelligence Without Reason, in ‘Proceedings, IJCAI-91’, Morgan Kaufmann Publishers, Inc., Sydney, Australia, pp. 569–595.
- Byrne, R. W. & Russon, A. E. (1998), ‘Learning by Imitation: a Hierarchical Approach’, *The Journal of Behavioral and Brain Sciences* **16**, 3.
- Daug, R., Blische, K. & Olivier, N. (1978), Scanning Habits and Visuo–Motor Learning, in J. K. O’Regan & A. Levy-Schoen, eds, ‘Eye Movements, From Physiology to Cognition: Selected/Edited Proceedings of the Third European Conference on Eye Movements’, pp. 323–332.
- Dautenhahn, K. (1995), ‘Getting to Know Each Other - Artificial Social Intelligence for Autonomous Robots’, *Robotics and Autonomous Systems* **16**, 333–356.
- Davis, J. M. (1973), Imitation: A Review and Critique, in Bateson & Klopfer, eds, ‘Perspectives in Ethology’, Vol. 1, Plenum Press, pp. 43–72.
- Decety, J. (1996), ‘Do imagined and executed actions share the same neural substrate?’, *Cognitive Brain Research* **3**, 87–93.
- Demiris, J. & Hayes, G. (1997), Do Robots Ape?, in ‘Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents (Technical Report FS-97-02)’, AAAI Press, Cambridge, Mass, pp. 28–30.
- Demiris, J., Rougeaux, S., Hayes, G. M., Berthouze, L. & Kuniyoshi, Y. (1997), Deferred Imitation of Human Head Movements by an Active Stereo Vision Head, in ‘Proceedings of the 6th IEEE International Workshop on Robot Human Communication (RoMan97)’, IEEE Press, Sendai, Japan.
- di Pellegrino, G., Fadiga, L., Fogassi, L. & Rizzolatti, G. (1992), ‘Understanding motor events: a neurophysiological study’, *Experimental Brain Research* **91**, 176–180.
- Donald, M. (1993), ‘Precis of Origins of the modern mind: Three stages in the evolution of culture and cognition’, *The Journal of Behavioral and Brain Sciences* **16**, 737–791.
- Elkins, J. (1996), *The object stares back: On the nature of seeing*, Simon & Schuster, NY.
- Essa, I. A. (1999), ‘Computers Seeing People’, *AI Magazine* **20**(2), 69–82.
- Fadiga, L., Fogassi, L., Pavesi, G. & Rizzolatti, G. (1995), ‘Motor facilitation during action observation: a magnetic stimulation study’, *J. Neurophysiol.* **73**(6), 2608–2611.
- Flanders, M. & Soechting, J. F. (1995), ‘Frames of Reference for Hand Orientation’, *Journal of Cognitive Neuroscience* **7**(2), 182–195.

- Flanders, M., Tillery, S. I. H. & Soechting, J. F. (1992), ‘Early stages in a sensorimotor transformation’, *Behavioral and Brain Sciences* **15**, 309–362.
- Gavrila, D. & Davis, L. (1995), ‘Towards 3-D Model-based Tracking and Recognition of Human Movement: a Multi-View Approach’, *International Workshop on the Face and Gesture Recognition*.
- Giszter, S. F., Mussa-Ivaldi, F. A. & Bizzi, E. (1993), ‘Convergent force fields organized in the frog’s spinal cord’, *Journal of Neuroscience* **13**(2), 467–491.
- Goodale, M. A., Jakobson, L. S., Milner, A. D., Perrett, D. I., Benson, P. J. & Hietanen, J. K. (1994), ‘The Nature and Limits of Orientation and Pattern Processing Supporting Visuomotor Control in a Visual Form Agnostic’, *Journal of Cognitive Neuroscience* **6**(1), 46–56.
- Hayes, G. & Demiris, J. (1994), A Robot Controller Using Learning by Imitation, in A. Borkowski & J. L. Crowley, eds, ‘Proceedings of the International Symposium on Intelligent Robotic Systems’, LIFIA-IMAG, Grenoble, France, pp. 198–204.
- Hogan, N. (1985), ‘Impedance Control: An Approach to Manipulation’, *Journal of Dynamic Systems, Measurement, and Control* **107**, 1–24.
- Hovland, G. E., Sikka, P. & McCarragher, B. J. (1996), Skill Acquisition from Human Demonstration Using a Hidden Markov Model, in ‘Proceedings, IEEE International Conference on Robotics and Automation’, Minneapolis, MN, pp. 2706–2711.
- Hubel, D. & Wiesel, T. (1977), ‘Functional Architecture of Macaque Monkey Cortex’, *Proceedings of the Royal Society of London* **198**, 1–59.
- Ikeuchi, K., Kawade, M. & Suehiro, T. (1993), Assembly Task Recognition with Planar, Curved, and Mechanical Contacts, in ‘Proceedings of IEEE International Conference on Robotics and Automation’, Atlanta, GA.
- Ikeuchi, K., Suehiro, T., Tanguy, P. & Wheeler, M. (1990), Assembly Plan from Observation, Technical report, Carnegie Mellon University Robotics Institute Annual Research Review.
- Jeannerod, M. & Decety, J. (1995), ‘Mental motor imagery: a window into the representational stages of action’, *Current Opinion in Neurobiology* **5**, 727–732.
- Jeannerod, M., Arbib, M. A., Rizzolatti, G. & Sakata, H. (1995), ‘Grasping objects: the cortical mechanisms of visuomotor transformation’, *Trends in Neuroscience* **18**, 314–320.
- Johansson, G. (1977), ‘Studies on visual perception of locomotion’, *Perception* **6**, 365–376.
- Kaiser, M. (1997), ‘Transfer of Elementary Skills via Human-Robot Interaction’, *Adaptive Behavior*.
- Kolb, B. & Milner, B. (1981), ‘Performance of Complex Arm and Facial Movements After Focal Brain Lesions’, *Neuropsychologia* **19**, 491–504.
- Kudo, H., Uomori, K., Yamada, N., Oinishi, N. & Sugie, N. (1992), Binocular Fixation-Point Shifts Induced by a Limb Occulusion, in ‘Proceedings of the IEEE/EMBS’, pp. 1670–1671.
- Kuniyoshi, Y., Inaba, M. & Inoue, H. (1994), ‘Learning by watching: extracting reusable task knowledge from visual observation of human performance’, *IEEE Transactions on Robotics and Automation* **10**(6), 799–822.

- Lee, C. & Xu, Y. (1996), Online, Interactive Learning of Gestures for Human/Robot Interfaces, *in* ‘Proceedings, IEEE International Conference on Robotics and Automation’, Vol. 4, Minneapolis, MN, pp. 2982–2987.
- Levine, I. N. (1983), *Quantum Chemistry*, Allyn and Bacon, Inc., Boston, MA.
- Marjanović, M., Scassellati, B. & Williamson, M. (1996), Self-Taught Visually-Guided Pointing for a Humanoid Robot, *in* P. Maes, M. Matarić, J.-A. Meyer, J. Pollack & S. Wilson, eds, ‘Fourth International Conference on Simulation of Adaptive Behavior’, The MIT Press, Cape Cod, MA, pp. 35–44.
- Martinetz, T. M. & Schulten, K. J. (1991), A neural-gas network learns topologies, *in* T. Kohonen, K. Mkišara, O. Simula & J. Kangas, eds, ‘Artificial Neural Networks’, North-Holland, Amsterdam, pp. 397–402.
- Matarić, M. J. (1994), Learning Motor Skills by Imitation, *in* ‘Proceedings, AAAI Spring Symposium Toward Physical Interaction and Manipulation’, Stanford University.
- Matarić, M. J. (1995), ‘Designing and Understanding Adaptive Group Behavior’, *Adaptive Behavior* **4**(1), 50–81.
- Matarić, M. J. (1997), ‘Behavior-Based Control: Examples from Navigation, Learning, and Group Behavior’, *Journal of Experimental and Theoretical Artificial Intelligence* **9**(2–3), 323–336.
- Matarić, M. J. & Pomplun, M. (1998), ‘Fixation Behavior in Observation and Imitation of Human Movement’, *to appear in Cognitive Brain Research*.
- Matarić, M. J., Zordan, V. B. & Williamson, M. (1999), ‘Making Complex Articulated Agents Dance: an analysis of control methods drawn from robotics, animation, and biology’, *Autonomous Agents and Multi-Agent Systems* **2**(1), 23–44.
- Meltzoff, A. N. (1993), Molyneux’s babies: Cross-modal perception, imitation and the mind of the preverbal infant, *in* N. Eilan, R. McCarthy & B. Brewer, eds, ‘Spatial Representation; Problems in Philosophy and Psychology’, Blackwell, pp. 219–235.
- Meltzoff, A. N. & Moore, M. K. (1983), ‘Newborn Infants Imitate Adult Facial Gestures’, *Child Development* **54**, 702–709.
- Meltzoff, A. N. & Moore, M. K. (1988), ‘Imitation of Facial and Manual Gestures by Human Neonates’, *Science* **198**, 75–78.
- Meltzoff, A. N. & Moore, M. K. (1994), ‘Imitation, Memory, and the Representation of Persons’, *Infant Behavior and Development* **17**, 83–99.
- Meltzoff, A. N. & Moore, M. K. (1995), Infants’ Understanding of People and Things: From Body Imitation to Folk Psychology, *in* J. L. Bermudez, A. Marcel & N. Eilan, eds, ‘The Body and the Self’, MIT Press/Bradford Books, pp. 44–69.
- Meltzoff, A. N. & Moore, M. K. (1997), ‘Explaining Facial Imitation: A Theoretical Model’, *Early Development and Parenting* **6**(2), 157.1–14.
- Moore, B. R. (1992), ‘Avian Movement Imitation and a new Form of Mimicry: Tracing the Evolution of a Complex Form of Learning’, *Behavior* **122**, 614–623.

- Moore, B. R. (1996), The Evolution of Imitative Learning, in C. M. Heyes & B. G. Galef, eds, 'Social Learning in Animals: The Roots of Culture', Academic Press, New York, pp. 245–265.
- Mussa-Ivaldi, F. A. & Giszter, S. (1992), 'Vector field approximation: a computational paradigm for motor control and learning', *Biological Cybernetics* **67**, 491–500.
- Mussa-Ivaldi, F. A., Giszter, S. F. & Bizzi, E. (1994), 'Linear combinations of primitives in vertebrate motor control', *Proceedings of the National Academy of Sciences* **91**, 7534–7538.
- Pelisson, D., Goodale, M. A. & Prablanc, C. (1978), Adjustments of Hand Pointings to Visual Targets do not Need Visual Reafference From the Moving Limb, in J. K. O'Regan & A. Levy-Schoen, eds, 'Eye Movements, From Physiology to Cognition: Selected/Edited Proceedings of the Third European Conference on Eye Movements', pp. 115–121.
- Perrett, D. I., Harries, M. H., Bevan, R., Thomas, S., Benson, P. J., Mistlin, A. J., Chitty, A. J., Hietanen, J. K. & Ortega, J. E. (1989a), 'Frameworks of Analysis for the Neural Representation of Animate Objects and Actions', *Journal of Experimental Biology* **146**, 87–113.
- Perrett, D. I., Harries, M. H., Mistlin, A. J. & Chitty, A. J. (1989b), Recognition of objects and actions: frameworks for neuronal computation and perceptual experience, in O. Guthrie, ed., 'Higher Order Sensory Processing', Manchester University Press.
- Perrett, D. I., Harries, M. H., Mistlin, A. J., Hietanen, J. K., Benson, P. J., Bevan, R., Thomas, S., Oram, M. W., Ortega, J. & Brierley, K. (1990), 'Social Signals Analyzed at the Single Cell Level: Someone is Looking at Me, Something Touched Me, Something Moved!', *International Journal of Comparative Psychology* **4**(1), 25–55.
- Perrett, D. I., Smith, P. A. J., Mistlin, A. J., Chitty, A. J., Head, A. S., Potter, D. D., Broennimann, R., Milner, A. D. & Jeeves, M. A. (1985), 'Visual Analysis of Body Movements by Neurones in the Temporal Cortex of the Macaque Monkey: A preliminary Report', *Behavioral Brain Research* **16**, 153–170.
- Piaget, J. (1962), *Play, Dreams and Imitation in Children*, W. W. Norton & Co., New York.
- Pook, P. K. & Ballard, D. H. (1993), Recognizing teleoperated manipulations, in 'Proceedings of IEEE International Conference on Robotics and Automation', Vol. 2, Atlanta, Georgia, pp. 578–585.
- Rabiner, L. R. (1989), 'A tutorial on hidden Markov models and selected applications in speech recognition', *Proceedings of the IEEE* **77**(2), 257–286.
- Rabiner, L. R. & Juang, B. H. (1986), 'Introduction to Hidden Markov Models', *IEEE ASSP Magazine* pp. 4–16.
- Rizzolatti, G., Fadiga, L., Gallese, V. & Fogassi, L. (1996), 'Premotor cortex and the recognition of motor actions', *Cognitive Brain Research* **3**, 131–141.
- Romann, J., F. E. (1999), 'Projective Virtual Reality: Bridging the Gap between Virtual Reality and Robotics', *IEEE Transaction on Robotics and Automation; Special Section on Virtual Reality in Robotics and Automation* **15**(3), 411–422.
- Savelsbergh, G. J., Whiting, H. T. A., Pijpers, J. R. & van Santvoord, A. A. M. (1993), 'The visual guidance of catching', *Experimental Brain Research* **93**, 148–156.

- Savelsvertgh, G. J. P. & Bootsma, R. J. (1994), 'Perception-action coupling in hitting and catching', *International Journal of Sport Psychology* **25**, 331–343.
- Schaal, S. (1997), Learning from demonstration, *in* M. Mozer, M. Jordan & T. Petsche, eds, 'Advances in Neural Information Processing Systems 9', The MIT Press, pp. 1040–1046.
- Stein, J. F. (1992), 'The representation of egocentric space in the posterior parietal cortex', *Behavior and Brain Science* **15**, 691–700.
- Tomasello, M., Kruger, A. C. & Rother, H. H. (1993), 'Cultural Learning', *The Journal of Behavioral and Brain Sciences* **16**(3), 495–552.
- Vogt, S. (1995), 'Imagery and perception-action mediation in imitative actions', *Cognitive Brain Research* **3**, 79–86.
- Williamson, M. (1996), Postural Primitives: Interactive Behavior for a Humanoid Robot Arm, *in* P. Maes, M. Mataric, J.-A. Meyer, J. Pollack & S. Wilson, eds, 'Fourth International Conference on Simulation of Adaptive Behavior', The MIT Press, Cape Cod, MA, pp. 124–131.
- Yang, J., Xu, Y. & Chen, C. S. (1997), 'Human Action Learning via Hidden Markov Model', *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans* **27**(1), 34–44.