Computational Approaches to Motor Learning by Imitation

Stefan Schaal[1,2], Auke Ijspeert[1,3], & Aude Billard[1,4]

[1]Computer Science & Neuroscience, University of Southern California,
3641 Watt Way, Los Angeles, CA 90089-2520

[2]ATR Human Information Sciences, 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0218, Japan

[3]School of Computer and Communication Sciences,

[4]School of Engineering Swiss Federal Institute of Technology,
Lausanne, CH 1015 Lausanne, Switzerland

Abstract

Movement imitation requires a complex set of mechanisms that map an observed movement of a teacher onto one's own movement apparatus. Relevant problems include movement recognition, pose estimation, pose tracking, body correspondence, coordinate transformation from external to egocentric space, matching of observed against previously learned movement, resolution of redundant degrees-of-freedom that are unconstraint by the observation, suitable movement representations for imitation, modularization of motor control, etc. All these topics by themselves are active research problems in computational and neurobiological sciences, such that their combination into a complete imitation system remains a daunting undertaking—indeed, one could argue that we need to understand the complete perception-action loop. As a strategy to untangle the complexity of imitation, this paper will examine imitation purely from a computational point of view, i.e., we will review statistical and mathematical approaches that have been suggested for tackling parts of the imitation problem, and discuss their merits, disadvantages, and underlying principles. Given the focus on action recognition of other contributions in this special issue, this paper will primarily emphasize the motor side of imitation, assuming that a perceptual system has already identified important features of a demonstrated movement and created their corresponding spatial information. Based on the formalization of motor control in terms of control policies and their associated performance criteria, useful taxonomies of imitation learning can be generated that clarify different approaches and future research directions.

Keywords: imitation, motor control, duality of movement generation and movement recognition, motor primitives

1   Introduction

Movement imitation is familiar to everybody from daily experience: a teacher demonstrates[1] a movement, and immediately the student is capable of approximately repeating it. Besides a variety of social, cultural, and cognitive implications that the ability to imitate entails (cf. reviews in Byrne & Russon 1998; Dautenhahn & Nehaniv 2002; Meltzoff & Moore 1994; Piaget 1951; Rizzolatti & Arbib 1998; Tomasello et al. 1993), from the viewpoint of learning, a teacher's demonstration as the starting point of one's own learning can significantly speed up the learning process, as imitation usually drastically reduces the amount of trial-and-error that is needed to accomplish the movement goal

---

[1] For the purpose of this paper, only visually mediated imitation will be considered, although, at least in humans, verbal communication can supply important additional information.
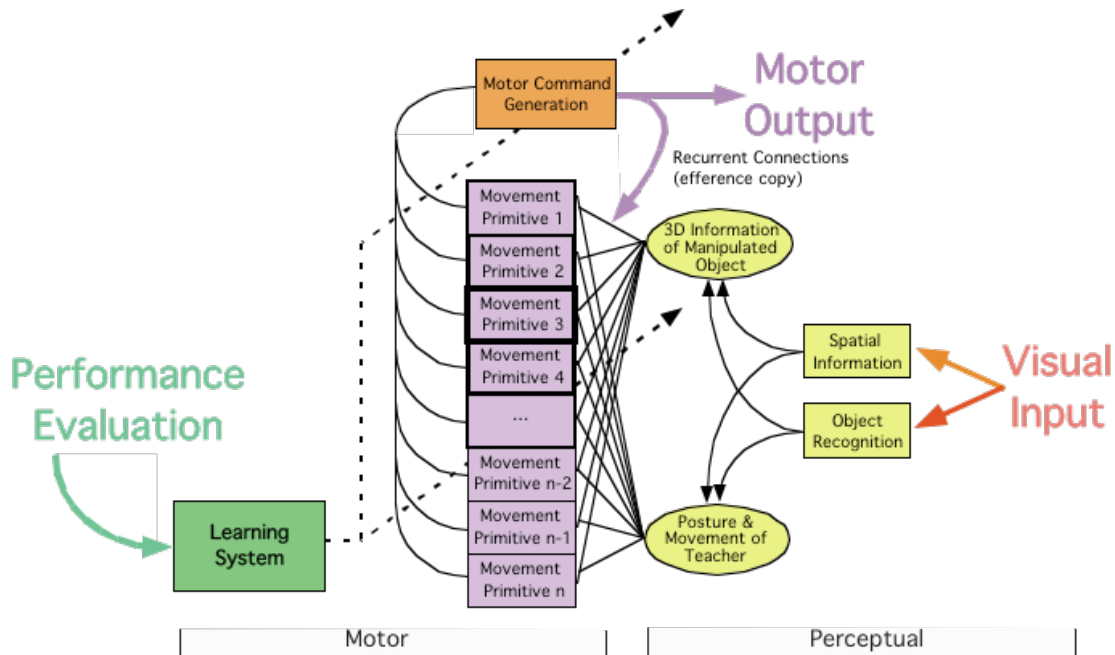
Figure 1: Conceptual sketch of an imitation learning system. The right side of the figure contains primarily perceptual elements and indicates how visual information is transformed into spatial and object information. The left side focuses on motor elements, illustrating how a set of movement primitives competes for a demonstrated behavior. Motor commands are generated from input of the most appropriate primitive. Learning can adjust both movement primitives and the motor command generator.

by providing a good example of a successful movement (Schaal 1999). Thus, from a computational point of view, it is important to understand the detailed principles, algorithms, and metrics that subserve imitation, starting from the visual perception of the teacher up to issuing motor commands that move the limbs of the student.

Figure 1 sketches the major ingredients of a conceptual imitation learning system (Schaal 1999). Visual sensory information needs to be parsed into information about objects and their spatial location in an internal or external coordinate system; the depicted organization is largely inspired by the dorsal (what) and ventral (where) stream as discovered in neuroscientific research (van Essen & Maunsell 1983). As a result, some form of postural information of the movement of the teacher and/or 3D object information about the manipulated object (if an object is involved) should become available. Subsequently, one of the major questions revolves around how such information can be converted into action. For this purpose, Figure 1 alludes to the concept of movement primitives, also called 'movement schemas', 'basis behaviors', 'units of action', 'macro actions', etc. (e.g., Arbib 1981; Dautenhahn & Nehaniv 2002; Sternad & Schaal 1999; Sutton et al. 1999). Movement primitives are sequences of action that accomplish a complete goal-directed behavior. They could be as simple as an elementary action of an actuator, e.g., 'go forward', go backward', etc., but, as discussed in Schaal (1999), such low-level representations do not scale well to learning in systems with many degrees-of-freedom. Thus, it is useful for a movement primitive to code complete temporal behaviors, like 'grasping a cup', 'walking', 'a tennis serve', etc. Figure 1 assumes that the

2

perceived action of the teacher is mapped onto a set of existing primitives in an assimilation phase, also suggested in Demiris and Hayes (2002) and (Wolpert et al. submitted). This mapping process also needs to resolve the correspondence problem concerning a mismatch between the teachers body and the student's body (Dautenhahn & Nehaniv 2002). Subsequently, the most appropriate primitives are adjusted by learning to improve the performance in an accommodation phase. Figure 1 indicates such a process by highlighting the better-matching primitives with increasing line widths. If no existing primitive is a good match for the observed behavior, a new primitive must be generated. After an initial imitation phase, self-improvement, e.g., with the help of a reinforcement-based performance evaluation criterion (Sutton & Barto 1998), can refine both movement primitives and an assumed stage of motor command generation (see below) until a desired level of motor performance is achieved.

In the following sections, we will attempt to formalize the conceptual picture of Figure 1 in the context of previous work on computational approaches to imitation. Given that Rittscher and Blake (submitted) already concentrate on the perceptual part of imitation in this issue, our review will focus on the motor side in Figure 1.

2   Computational Imitation Learning

Initially, at the beginning of the 1980ies, computational imitation learning found the strongest research interest in the field of manipulator robotics, as it seemed to be a promising route to automate the tedious manual programming of these machines. Inspired by the ideas of artificial intelligence, symbolic reasoning was the common choice to approach imitation, mostly by parsing a demonstrated movement into some form of "if-then" rules that, when chained together, created a finite state machine controller (e.g., Dufay & Latombe 1984; Levas & Selfridge 1984; Lozano-Pérez 1982; Segre & De-Jong 1985; Segre 1988). Given the reduced computational power available at this time, a demonstration normally consisted of manually "pushing" the robot through a movement sequence and using the proprioceptive information that the robot sensed during this guided movement as basis to extract the if-then rules. In essence, many recent robotics approaches to imitation learning have remained closely related to this strategy. New elements include the use of visual input from the teacher and movement segmentation derived from computer vision algorithms (Ikeuchi et al. 1993; Kuniyoshi et al. 1989; Kuniyoshi et al. 1994). Other projects used data gloves or marker-based observation systems as input for imitation learning (Tung & Kak 1995).

More recently, research on imitation learning has been influenced increasingly by non-symbolic learning tools, for instance artificial neural networks, fuzzy logic, statistical learning, etc. (Dillmann et al. 1995; Hovland et al. 1996; Pook & Ballard 1993). An even more recent trend takes inspiration of the known behavioral and neuroscientific processes of animal imitation to develop algorithms for robot programming by demonstration (e.g., Arbib et al. 2000; Billard 2000; Oztop & Arbib 2002) with the goal of developing a more general and less task specific theory of imitation learning. It is these neural computation techniques that we will focus on in this review, as they offer the most to both biologically inspired modeling of imitation and technological realizations of imitation in artificial intelligence systems.

2.1   A Computational Formalization of Imitation Learning

Successful motor control requires issuing motor commands for all the actuators of a movement system at the right time and of correct magnitude in response to internal and external sensations and a given behavioral goal. Thus, the problem of motor control can generally be formalized as finding a task-specific *control policy* π:

$$\mathbf{u}(t) = \pi(\mathbf{z}(t), t, \boldsymbol{\alpha}) \tag{1}$$

where **u** denotes the vector of motor commands, **z** the vector of all relevant internal states of the movement system and external states of the environment, t represents the time parameter, and $\boldsymbol{\alpha}$ stands for the vector of open parameters that need to be adjusted during learning, e.g., the weights of a neural network (Dyer & McReynolds 1970). We will denote a policy that explicitly uses a dependence on time as a *nonautonomous* policy, while a policy without explicit time dependence, i.e., $\mathbf{u}(t) = \pi(\mathbf{z}(t), \alpha)$, will be called *autonomous*. The formulation in (1) is very general and can be applied to any level of analysis, like a detailed neuronal level or a more abstract joint angular level. If the function π were known, the task goal could be achieved from every state **z** of the movement system. This theoretical view allows us to reformulate imitation learning in terms of the more formal question of how control policies, which we also call movement primitives, can be learned (or bootstrapped) by watching a demonstration.

Crucial to the issue of imitation is a second formal element, an evaluation criterion that creates a metric of the level of success of imitation:

$$J = g(\mathbf{z}(t), \mathbf{u}(t), t) \tag{2}$$

Without any loss of generality, we will assume that the cost J should be minimized; particular instantiations of J will be discussed below. In general, J can be any kind of cost function, defined as an accumulative cost over a longer time horizon as is needed for minimizing energy, or only over one instant of time, e.g., as needed when trying to reach a particular goal state. Moreover, J may be defined on variables based in any coordinate system, e.g., external, internal, or a mixed set of coordinates. The different ways of creating control policies and metrics will prove to be a useful taxonomy of previous approaches to imitation learning and the problem of imitation in general.

Defining the cost J for an imitation task is a complex problem. In an ideal scenario, J should capture the task goal and the quality of imitation in achieving the task goal. For instance, the task goal could be to reach for a cup, which could be formalized as a cost that penalizes the squared distance between the hand and the cup. The teacher's demonstration, however, may have chosen a particular form of reaching for the cup, e.g., in a strangely curved hand trajectory. Thus, faithful imitation may require adding an additional term to the cost J that penalizes deviations from the trajectory the teacher demonstrated, depending on whether the objective of imitation is solely focused on the task or also on how to move in order to perform the task. Hence, the cost J quickly becomes a complex, hybrid criterion defined over various objectives. In biological research, it often remains a hard problem to discover what kind of metric the student applied when imitating (Mataric & Pomplun 1998; Nehaniv & Dautenhahn 1999).
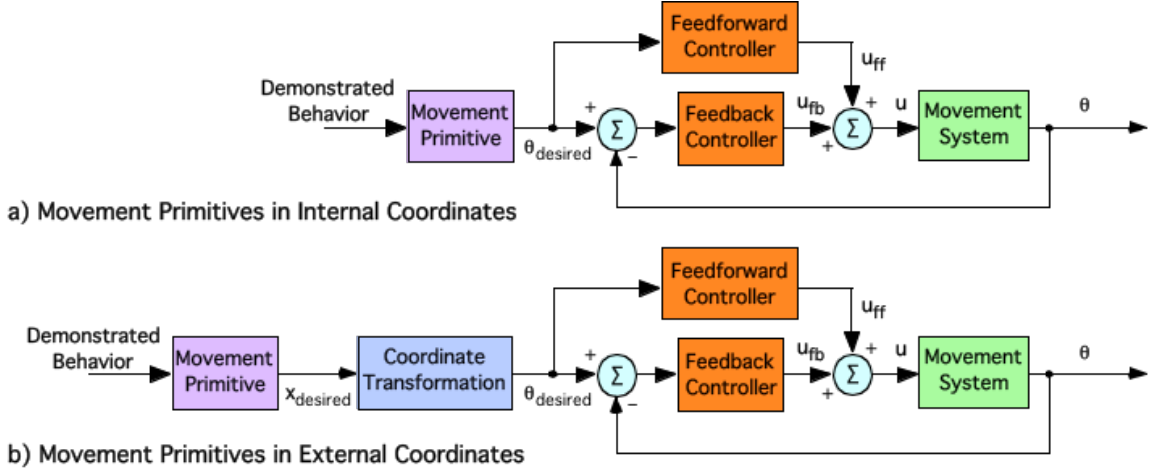
Figure 2: Modular motor control with movement primitives, using a) a movement primitive defined in internal coordinates, and b) a movement primitive defined in external coordinates.

2.2 Imitation by Direct Policy Learning

The demonstrated behavior can be used to learn the appropriate control policy directly by supervised learning of the parameters α of the policy (cf. Equation (1)), i.e., a nonlinear map $\mathbf{z} \rightarrow \mathbf{u}$, employing an *autonomous* policy and using as evaluation criterion (cf. Equation (2)) simply the squared error of reproducing $\mathbf{u}$ in a given state $\mathbf{z}$. For this purpose, the state $\mathbf{z}$ and the action $\mathbf{u}$ of the teacher need to be observable and identifiable, and they must be meaningful for the student, i.e., match the student's kinematic and dynamic structure (cf. Dautenhahn & Nehaniv 2002). This prerequisite of observability, shared by all forms of imitation learning, imposes a serious constraint since, normally, motor commands, i.e., kinetic variables, and internal variables of the teacher are hidden from the observer. While statistical learning has methods to uncover hidden states, e.g., by means of Hidden Markov Models, Kalman filters, or more advanced methods (Arulampalam et al. 2002), we are not aware that such techniques have been applied to imitation yet.

Thus, in order to instantiate a movement primitive from a demonstration, the primitive needs to be defined in variables that can be perceived, leaving only kinematic variables as potential candidates, e.g., positions, velocities, and accelerations. Given that the output of a movement primitive has to be interpreted as some form of a command to the motor system, usually implying a desired change of state, movement primitives that output a desired velocity or acceleration may be useful, i.e., a "desired time-derivative" of the state information[2] that is used to represent the teacher's movement. Our generic formulation of a policy in Equation (1) may hence be written more suitably as

$$\dot{\mathbf{z}}(t) = \pi\big(\mathbf{z}(t), t, \alpha\big) \tag{3}$$

From a control theoretic point of view, this line of reasoning requires that motor control be modular, i.e., has at least separate processes for movement planning (i.e., generating

---

[2] Note that instead of a formulation in terms of differential equation, we would also choose difference equation, i.e., where a desired "next state" is the output of the policy, not a desired change of state.

the right kinematics) and execution (i.e., generating the right dynamics) (Wolpert 1997; Wolpert & Kawato 1998). Figure 2 illustrates two classical examples (e.g., Craig 1986) of modular control in the context of imitation learning and motor primitives. In Figure 2a, the demonstrated behavior is mapped onto a movement primitive that is defined in internal coordinates of the student – joint angular coordinates θ are a good candidate as they can be extracted from visual information, a problem addressed under the name of pose estimation in computer vision (Deutscher et al. 2000; Rittscher & Blake submitted). Such internal coordinates can directly serve as desired input to a motor command execution stage (cf. Figure 1), here assumed to be composed of a feedback and a feedforward control block (Kawato 1999).

Alternatively, Figure 2b illustrates the subtle but important change when movement primitives are represented in external coordinates, i.e., a task-level representation (Aboaf et al. 1989; Saltzman & Kelso 1987). For instance, the acceleration of the fingertip in the task of pole balancing would be interpreted as a task-level command issued by the movement primitive in external coordinates, in contrast to joint angular accelerations of the entire arm and body that would be issued by a movement primitive in internal coordinates. Most often, task-level representations are easier to extract from a demonstration, and have a more compact representation. Task-level representations can also cope with a mismatch in dynamic and/or kinematic structure between the teacher and the student—only the task state is represented, not the state of motor system that performs the task. Task-level imitation requires prior knowledge of how a task-level command can be converted into a command in internal coordinates, a classic problem in control theory treated under the name of inverse kinematics (Baillieul & Martin 1990), but which has found several elegant solutions in neural computation in the recent years (Bullock et al. 1993; D'Souza et al. 2001; Guenther & Barreca 1997).

In summary, movement primitives for imitation learning seem to be the most useful if expressed in kinematic coordinates, either in internal (e.g., joint, muscle) space

$$\dot{\theta}(t) = \pi(\mathbf{z}(t), t, \boldsymbol{\alpha}) \tag{4}$$

or in external (task) space

$$\dot{\mathbf{x}}(t) = \pi(\mathbf{z}(t), t, \boldsymbol{\alpha}) \tag{5}$$

Note that the formulations in Equation (4) and (5) intentionally use $\mathbf{z}$, the variable that represents all possible state information about the movement system and the environment as input, but only output a variable that is the desired change of state of the student in the selected coordinate system, i.e., $\dot{\mathbf{x}}$ in external space, and $\dot{\theta}$ in internal space. By dropping the explicit time dependence on the right side of (4) and (5), both policy formulations can be made autonomous.

Direct policy learning from imitation can now be reviewed more precisely in the context of the discussions of the previous paragraphs and Figure 2. Direct policy learning in task space was conducted for the task of pole balancing with a computer-simulated pole (Nechyba & Xu 1995; Widrow & Smith 1964). For this purpose, a supervised neural network was trained on task-level data recorded from a human demonstration. Similarly, several mobile robotics groups adopted imitation by direct policy learning using a "robot teacher" (Dautenhahn 1995; Grudic & Lawrence 1996; Hayes &

Demiris 1994; Lin 1991). For example, the "robot student" followed the "robot teacher's" movements in a specific environment, mimicked its kinematic, task-oriented actions, and learned to associate which action to choose in which state. Afterwards, the robot student had the same competence as the teacher in this environment. An impressive application of direct policy learning in a rather complex control system, a flight simulator, was demonstrated by (Sammut et al. 1992). Kinematic control actions from several human subjects were recorded and an inductive machine learning algorithm was trained to represent the control with decision trees. Subsequently, the system was able to autonomously perform various flight maneuvers.

In all these direct policy-learning approaches, there is no need for the student to know the task goal of the teacher, i.e., Equation (2) has only imitation-specific criteria, but no task-specific criteria. Imitation learning is greatly simplified in this manner. However, the student will not be able to undergo self-improvement unless an explicit reward signal, usually generated from a task-specific optimization criterion, is provided to the student, as in approaches discussed below. Another problem with direct policy learning is that there is no guarantee that the imitated behavior is stable, i.e., can reach the (implicit) behavioral goal from all start configurations. Lastly, imitation by direct policy learning usually generates policies that cannot be re-used for a slightly modified behavioral goal. For instance, if reaching for a specific target was learned by direct policy learning, and the target location changes, the commands issued by the learned policy are wrong for the new target location. Such a form of imitation of is often called "indiscriminate imitation" or "mimicking" as it just repeats an observed action pattern without knowledge about how to modify it for a new behavioral context.

2.3   Imitation by Learning Policies from Demonstrated Trajectories

A teacher's demonstration usually provides a rather limited amount of data, best described as "sample trajectories". Various projects investigated how a stable policy can be instantiated from such small amount of information. As a crucial difference with respect to direct policy learning, it is now assumed that the task goal is known (see examples below), and the demonstrated movement is only used as a seed for an initial policy, to be optimized by a self-improvement process. This self-learning adjusts the imitated movement to kinematic and dynamic discrepancies between the student and the teacher, and additionally ensures behavioral stability.

The idea of learning from trajectories was explored with an anthropomorphic robot arm for dynamic manipulation tasks, for instance, learning a tennis forehand and the game of kendama ('ball-in-the-cup') (Miyamoto & Kawato 1998; Miyamoto et al. 1996). At the outset, a human demonstrated the task, and his/her movement trajectory was recorded with marker-based optical recording equipment (OptoTrack). This process resulted in spatio-temporal data about the movement of the manipulated object in Cartesian coordinates, as well as the movement of the actuator (arm) in terms of joint angle coordinates. For imitation learning, a hybrid internal/external evaluation criterion was chosen. Initially, the robot aimed at indiscriminate imitation of the demonstrated trajectory in task space based on position data of the endeffector, while trying to use an arm posture as similar as possible to the demonstrated posture of the teacher (cf. D'Souza et al. 2001). This approximation process corrected for kinematic differences

between the teacher and the robot and resulted in a desired trajectory for the robot's motion—a desired trajectory can also be conceived of as a nonautonomous policy (Schaal et al. 2000). Afterwards, using manually provided knowledge of the task goal in form of an optimization criterion, the robot's performance improved by trial and error learning until the task was accomplished. For this purpose, the desired endeffector trajectory of the robot was approximated by splines, and the spline nodes, called viapoints, were adjusted in space and time by optimization techniques (e.g., Dyer & McReynolds 1970) until the task was fulfilled. Using this method, the robot learned to manipulate a stochastic, dynamic environment within a few trials.

A spline-based encoding of a control policy is nonautonomous, since the viapoints defining the splines are parameterized explicitly in time. There are two drawbacks in using such nonautonomous movement primitives. First, modifying the policy for a different behavioral context, e.g., a change of target in reaching or a change of timing and amplitude in a locomotion pattern, requires more complex computations in terms of scaling laws of the via points (Kawamura & Fukao 1994). Second, and more severely, nonautonomous policies are not very robust in coping with unforeseen perturbations of the movement. For instance, when abruptly holding the arm of a tennis player during a forehand swing, a nonautonomous policy would continue creating desired values for the movement system, and, due to the explicit time dependency, these desired values would increasingly more open a large gap between the current position and the desired position. This gap can potentially cause huge motor commands that fight the advert perturbation, and, if the arm were released, it would "jump" to catch up with the target trajectory, a behavior that is undesirable in any motor system as it leads to potential damage. In contrast, autonomous movement primitives can avoid this behavior as the output of the policy is solely state and not time dependent, and perturbations can create inhibitive terms in the policy that ensure that the planned movement of the policy will never deviate too much from the actual position. In this vein, Ijspeert, Nakanishi, and Schaal (Ijspeert et al. 2002a; Ijspeert et al. 2002b) suggested the use of autonomous dynamical systems as an alternative to spline-based imitation learning, realizing that Equations (4) and (5) are nothing but nonlinear differential equations. In their approach, a demonstrated trajectory is encoded by learning the transformation from a simple canonical attractor system to a new nonlinear attractor landscape that has the demonstrated trajectory as its unique attractor. Both limit cycle or point attractors could be realized, corresponding to rhythmic or discrete movement primitives. The evaluation criterion for imitation was the deviation of the reproduced trajectory from the demonstrated one, either in internal or external space—reaching the target of the movement, i.e., either a point or a limit cycle, is automatically guaranteed by shaping the attractor landscape appropriately. The dynamic systems policies were designed to provide spatial and temporal invariant, i.e., a qualitatively similar movement will always lead to a similarly parameterized movement primitive, irrespective of the timing of the movement and the target to which the movement was executed. Coupling terms to the differential equations allowed natural robustness towards external perturbations (see also Hatsopoulos 1996). The effectiveness of imitation learning with these dynamic systems primitives was successfully demonstrated on a humanoid robot that learned a series of movements such as tennis forehand, tennis backhand, and drumming sequences from a
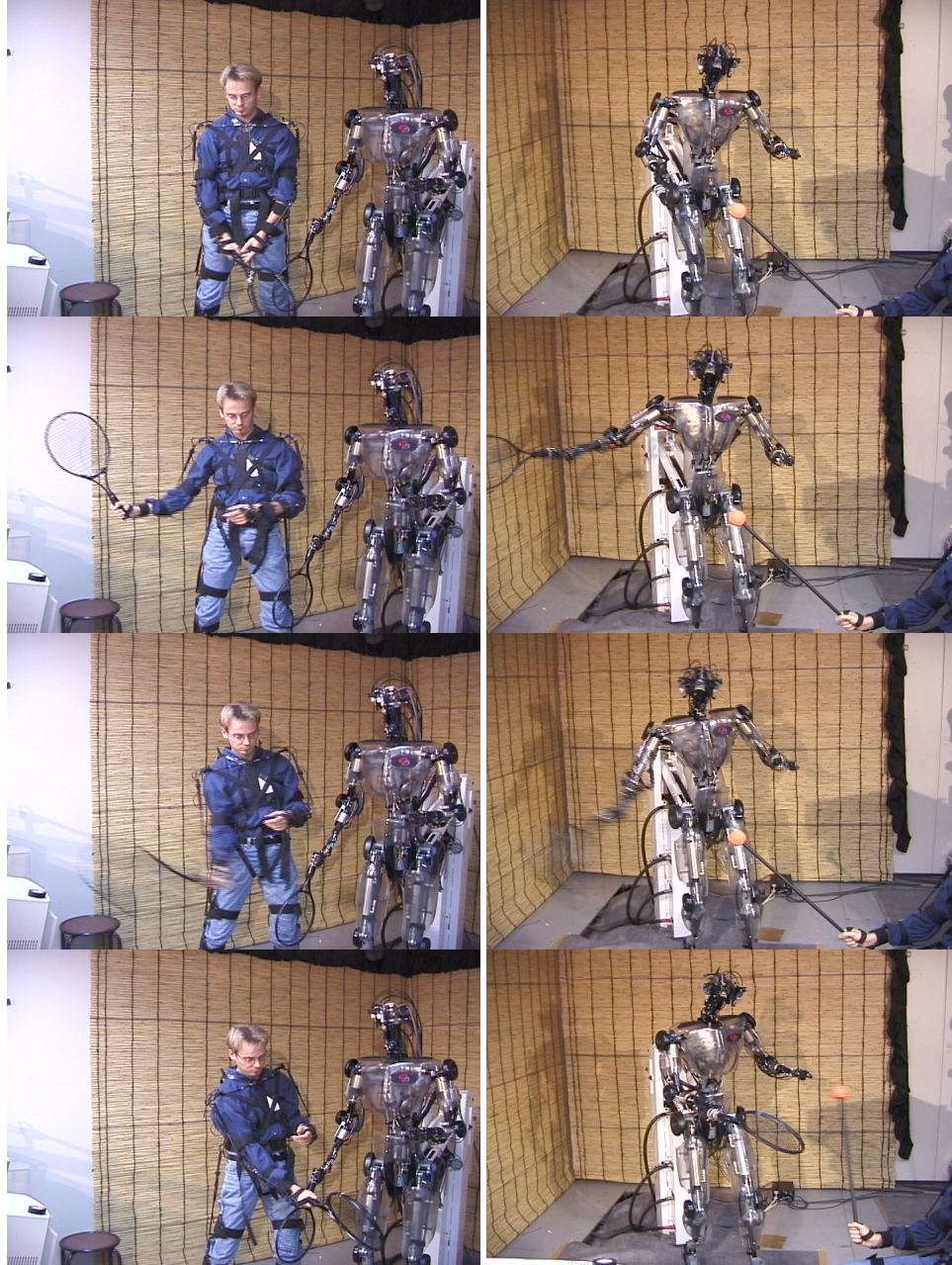
Figure 3: Left Column: Teacher demonstration of a tennis swing, Right Column: Imitated movement by the humanoid robot.

human teacher (Figure 3), and that was subsequently able to re-used the learned movement in modified behavioral contexts.

Another, more biologically inspired, dynamic system's approach to imitation was pursued by Billard et al. (Billard 2000; Billard & Mataric 2001; Billard & Schaal 2001). Joint angular trajectories, recorded from human demonstrations, were segmented using zero velocity points. The policy approximated the segment for each joint movement by a second order differential equation that activated a pair of antagonistic muscles, modeled as spring-damper systems (Lacquaniti & Soechting 1986). Due to the dynamic
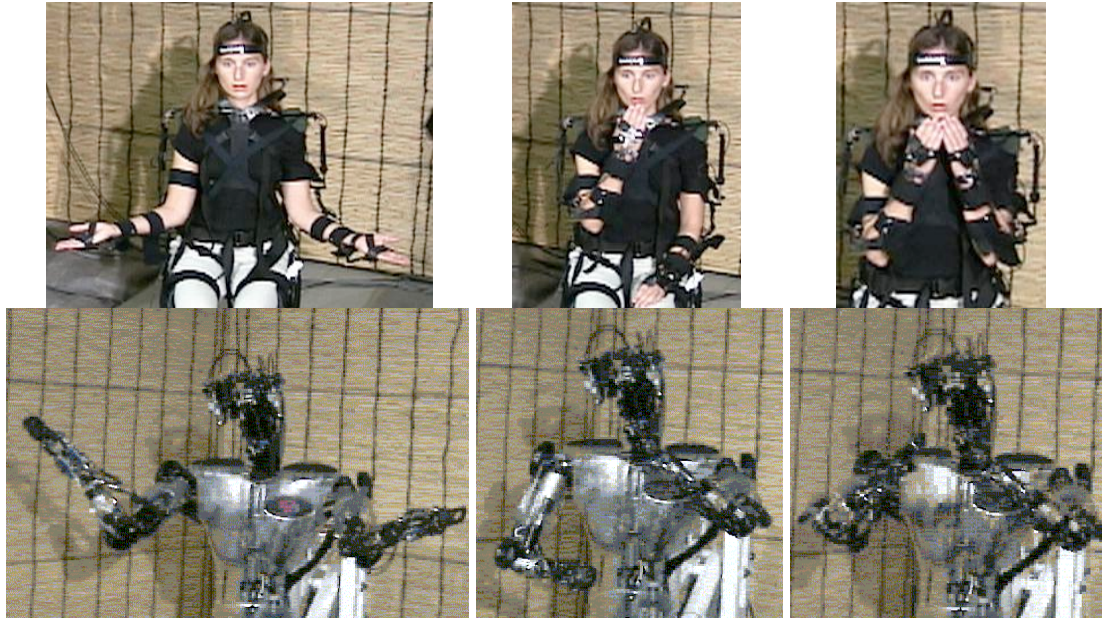
Figure 4: Learning of movement sequences by imitation.

properties of muscles, this policy generates joint angle trajectories with a bell-shape velocity profile similarly to human motion; the initial flexion or extension force determines entirely the trajectory and is computed using the initial acceleration of the demonstrated trajectory segment. After acquiring this movement primitive, imitation learning is used to combine joint trajectory segments to produce whole body motion. For this purpose, a time-delay recurrent neural network is trained to reproduce the sequential activation of each joint, similar to methods of associative memory (Schwenker et al. 1996). Both speed an amplitude of movement can be modulated by adjusting appropriate parameters in the network. This imitation system can generate complex movement sequences (Figure 4) and even "improvise" movement by randomly activating nodes in the associative memory.

2.4 Imitation by Model-based Policy Learning

A third approach to learning a policy from imitation employs model-based learning (Atkeson & Schaal 1997a; Schaal 1997). From the demonstrated behavior, not the policy but a predictive model of the task dynamics is approximated (cf. Wolpert et al. 1998). Given knowledge of the task goal, the task-level policy of the movement primitive can be computed with reinforcement learning procedures based on the learned model. For example, Schaal and Atkeson (Atkeson & Schaal 1997a; Atkeson & Schaal 1997b; Schaal 1997) showed how the model-based approach allowed an anthropomorphic robot arm to learn the task of pole-balancing in just a single trial, and the task of a "pendulum swing-up" in only three to four trials. These authors also demonstrated that task-level imitation based on direct policy learning, augmented with subsequent self-learning, can be rather fragile and does not necessarily provide significant learning speed improvement over pure trial-and-error learning without a demonstration

2.5 Matching of Demonstrated Behavior against Existing Movement Primitives

The approaches discussed in the previous paragraphs illustrated some computational ideas for how *novel* behaviors can be learned by imitation. Interesting insights into these methods can be gained by analyzing the process of how a perceived behavior is mapped onto a set of existing primitives. Two major questions (Meltzoff & Moore 1997) become a) what is the matching criterion for recognizing a behavior, and b) in which coordinate frame does matching take place?

Matching based on Policies with Kinetic Outputs

If only a kinetic control policy of the movement primitive exists (cf. Equation (1)), finding a matching criterion becomes difficult since kinetic outputs like forces or torques cannot be observed from demonstrations. One solution would be to execute a primitive, observe its outcome in either internal or external kinematic space, and generate in the chosen coordinate frame a performance criterion based on the similarity between the executed and the teacher's behavior, e.g., the squared difference of state variables over time or distance to a goal at the end of the movement. This procedure needs to be repeated for every primitive in the repertoire and is thus quite inefficient. Given that kinetic policies are also not very useful for learning novel movements by imitation (cf. Section 2.2), kinetic policies seem to be of little use in imitation learning.

Matching based on Policies with Kinematic Outputs

If the primitive outputs observable variables, e.g., kinematic commands as in Equations (4) and (5), matching is highly simplified since the output of the primitive can be compared directly with the teacher's performance. Such kinematic matching assumes that the motor execution stage of Figure 2 creates motor commands that faithfully realize the kinematic plans of the primitive, i.e., that motor command generation approximately inverts the dynamics of the movement system (Kawato 1999). At least two forms of matching mechanisms are possible.

One matching paradigm simply treats the demonstrated movement as a candidate for a *new* movement primitive and fits the parameterization of this primitive. The parameters are subsequently compared against the parameters of all previously learned primitives, and the best matching one in memory is chosen as the winner. For this method to work, the parameterization of the movement primitive should have suitable invariances towards variations of a movement, e.g., temporal and spatial scale invariance. The via-point method of Miyamoto et al. (1996) can easily be adapted for such movement recognition, as via-points represent a parsimonious parameterization of a movement that is easily used in classification algorithms, e.g., nearest neighbor methods (Wada & Kawato 1995). Similarly, the dynamic systems approach to motor primitives of Ijspeert et al. (Ijspeert et al. 2002b) creates a movement parameterization that affords classification in parameter space—indeed, the in-built scale and time invariances of this technique adds significant robustness to movement recognition in comparison to methods.

The second matching paradigm is based on the idea of predictive forward models (Atkeson & Schaal 1997a; Demiris & Hayes 2002; Miall & Wolpert 1996; Schaal 1997; Wolpert et al. submitted; Wolpert et al. 1998). While observing the teacher, each move-

ment primitive can try to predict the temporal evolution of the observed movement based on the current state $\mathbf{z}$ of the teacher. The primitive with the best prediction abilities will be selected as the best match. If, as mentioned above, the motor execution stage of the control circuit (Figure 2) faithfully realizes the movement plan issued by a movement primitive, the primitive can act itself as a forward model, i.e., it can predict a change in state $\mathbf{z}$ of the teacher (cf. Equations (4) and (5)). Alternatively, it is also possible to include prediction over the entire dynamics of the movement system. For this purpose, the output of the movement primitive is fed to the motor command execution stage, whose output is subsequently passed through a predictive forward model of the dynamics of the student's movement system (see Demiris & Hayes 2002; Wolpert et al. submitted in this issue), thus predicting the change of state of movement without actually performing it. This technique will work even when the motor execution stage is less accurate in realizing desired movement kinematics, but it comes at the cost of two more levels of signal processing, i.e., the simulated motor command generation and the need for a forward model of the motor system. Demiris and Hayes (Demiris & Hayes 2002) realized such an imitation system in a simulated humanoid.

What is particularly noteworthy in the above approaches to movement recognition is the suggested bidirectional interaction between perception and action: Movement recognition is directly accomplished with the movement generating mechanism. This concept is compatible with what the concept of mirror neurons in neurobiology (Rizzolatti & Arbib 1998; Rizzolatti et al. 1996), with the simulation theory of mind reading (Gallese & Goldman 1998), and it also ties into other research projects that emphasize the bidirectional interaction of generative and recognition models (Dayan et al. 1995; Kawato 1996) in unsupervised learning. Such bidirectional theories enjoy an increasing popularity in theoretical models to perception and action as they provide useful constraints for explaining the autonomous development of such system.

Matching based on Other Criteria

Exploiting the literature on computer vision and statistical classification, a large variety of alternative approaches to movement recognition can be developed, mostly without taking into account mutuality criteria between movement generation and movement recognition. Rittscher and Bake (submitted) in this issue provide an overview of techniques in this vein.

2.6 The Correspondence Problem

An important topic of imitation learning concerns how to map the external and internal space of the teacher to the student, often called the "correspondence problem" (Alissandrakis et al. 2002; Byrne submitted). Solving correspondence in external space is usually simplified, as external coordinates (or task coordinates) are mostly independent of the kinematic and dynamic structure of the teacher. For instance, if pole balancing could be demonstrated by a dolphin, a human student could imitate despite the mismatch in body structure if only task-level imitation is attempted—the only transformation needed is a mapping from the teacher's body-centered external space to the student's body-centered external space, which is just a linear transformation. Correspondence in internal space is a more complex problem. Even when teacher and student

have the same degrees of freedom, as it is the case with human-to-human or human-to-humanoid-robot imitation, the bodies of student and teacher are bound to differ in many ways, including in their ranges of motion, in their exact kinematics, and their dynamics. The mapping is even more difficult when the teacher and student have dissimilar bodies. In that case, the student can make only imitate approximately, reproducing only subgoals or sub-states of the demonstrated motion. The correspondence problem consists of defining which sub-states of the motion *can* and/or *should* be reproduced. Nehaniv and Dautenhahn (Dautenhahn & Nehaniv 2002) proposed a general mathematical framework to express such a mapping function in terms of transfer functions across different spaces. Alissandrakis et al. (Alissandrakis et al. 2002) implement this framework to solve the correspondence problem in a chess game case study. The movement of two chess pieces (e.g. queen and knight) are directed by very different rules such that the two pieces cannot replicate each other move in just one time step. In order for the knight to replicate the trajectory followed by the queen, it must define a number of subgoals (positions on the chessboard) through which the queen has traveled and that the knight can reach using its own movement capacities. The best strategy to define the subgoals depends on the metric applied to measure the imitation performance. The authors compare metrics that minimizes either the total number of moves required for the reproduction, or the space covered during the reproduction by the motion.

2.7   Imitation of Complex Movement Sequences

One final issue concerns the imitation of complex motor acts that involve learning a sequence of primitives and when to switch between them. In this context, Fagg and Arbib (1998) provided a model of reaching and grasping based on the known anatomy of the fronto-parietal circuits, including the mirror neuron system. Essentially, their model employed a recurrent neural network that sequenced and switched between motor schemas based on sensory cues. The work of Billard et al. (Section 2.3) follows a similar vein, just at a higher level of biological abstraction and more suitable for the control of real, complex robotic systems. In a robotic study, Pook and Ballard (1993) used hidden Markov models to learn appropriate sequencing from demonstrated behavior for a dexterous manipulation task. There is also large body of literature in the field of time series segmentation (Cacciatore & Nowlan 1994; Pawelzik et al. 1996; Weigend et al. 1995) that employed competitive learning and forward models for recognition and sequencing in a way that is easily adapted for imitation learning as illustrated in Figure 1.

3   Summary

Using the formalization of motor control in terms of generating control policies under a chosen performance criterion, we discussed computational imitation learning as methodology to bootstrap a student's control policy from a teacher's demonstration. Different methods of imitation were classified according to which variables were assumed observable for the student, whether variables were of kinetic or kinematic nature, whether internal, external coordinates, or both were used during demonstration, and whether the task goal was explicitly known to the student or not. Additional insights could be obtained by discussing how a demonstrated movement can be mapped onto a

set of existing movement primitives. Important topics in computational imitation concerned the formation of motor primitives, their representation, their sequencing, the reciprocal interaction of movement recognition and movement generation, and the correspondence problem. At the current stage of research, all these issues have been modeled in various ways, demonstrating an increasingly growing formal understanding of how imitation learning can be accomplished. Among the most crucial missing points to be addressed in imitation is presumably a formalization of extracting the intent of a demonstrated movement. Billard et al. (Billard & Schaal 2002) suggested some first ideas towards this goal by modeling probability distribution over manipulated objects by the teacher, which triggered appropriate imitation behavior in a humanoid robot. However, a more abstract representation of task goals, maybe as set of a generic goal taxonomy may be needed to make further progress in this area.

4   Acknowledgements

5   References

Aboaf, E. W., Drucker, S. M. & Atkeson, C. G. 1989 Task-level robot learing: Juggling a tennis ball more accurately. In *Proceedings of IEEE Interational Conference on Robotics and Automation*, pp. 331-348. May 14-19, Scottsdale, Arizona: Piscataway, NJ: IEEE.

Alissandrakis, A., Nehaniv, C. L. & Dautenhahn, K. 2002 Imitating with ALICE: Learning to imitate corresponding actions across dissimilar embodiments. *IEEE Transactions on Systems, Man, & Cybernetics, Part A: Systems and Humans* **32**, 482-496.

Arbib, M. A. 1981 Perceptual structures and distributed motor control. In *Handbook of Physiology, Section 2: The Nervous System Vol. II, Motor Control, Part 1* (ed. V. B. Brooks), pp. 1449-1480: Bethesda, MD: American Physiological Society.

Arbib, M. A., Billard, A., Iacoboni, M. & Oztop, E. 2000 Synthetic brain imaging: grasping, mirror neurons and imitation. *Neural Netw* **13**, 975-97.

Arulampalam, S., Maskell, S., Gordon, N. & Clapp, T. 2002 A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing* **50**, 174-188.

Atkeson, C. G. & Schaal, S. 1997a Learning tasks from a single demonstration. In *IEEE International Conference on Robotics and Automation (ICRA97)*, vol. 2, pp. 1706-1712. Albuquerque, NM, 20-25 April: Piscataway, NJ: IEEE.

Atkeson, C. G. & Schaal, S. 1997b Robot learning from demonstration. In *Machine Learning: Proceedings of the Fourteenth International Conference (ICML '97)* (ed. D. H. Fisher Jr.), pp. 12-20. Nashville, TN, July 8-12, 1997: Morgan Kaufmann.

Baillieul, J. & Martin, D. P. 1990 Resolution of kinematic redundancy. In *Proceedings of Symposia in Applied Mathematics*, vol. 41, pp. 49-89. San Diego, May 1990: Providence, RI: American Mathematical Society.

Billard, A. 2000 Learning motor skills by imitation: A biologically inspired robotic model. *Cybernetics and Systems* **32**, 155-193.

Billard, A. & Mataric, M. 2001 Learning human arm movements by imitation: Evaluation of a biologically-inspired architecture. *Robotics and Autonomous Systems* **941**, 1-16.

Billard, A. & Schaal, S. 2001 A connectionist model for on-line robot learning by imitation. In *IEEE International Conference on Intelligent Robots and Systems (IROS 2001)*. Maui, Hawaii, Oct.29-Nov.3: Piscataway, NJ: IEEE.

Billard, A. & Schaal, S. 2002 Computational elements of robot learning by imitation. In *American Mathematical Society Central Section Meeting*. Madison, Oct.12-13,2002: Providence, RI: American Mathematical Society.

Bullock, D., Grossberg, S. & Guenther, F. H. 1993 A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm. *Journal of Cognitive Neuroscience* **5**, 408-435.

Byrne, R. submitted Imitation as behavior parsing. *Philosophical Transactions of the Royal Society of London B*.

Byrne, R. W. & Russon, A. E. 1998 Learning by imitation: a hierarchical approach. *Behav Brain Sci* **21**, 667-84; discussion 684-721.

Cacciatore, T. W. & Nowlan, S. J. 1994 Mixtures of controllers for jump linear and non-linear plants. In *Advances in Neural Information Processign Systems 6* (ed. J. D. Cowen, G. Tesauro & J. Alspector), pp. 719-726. San Mateo, CA: Morgan Kaufmann.

Craig, J. J. 1986 *Introduction to robotics*. Reading, MA: Addison-Wesley.

D'Souza, A., Vijayakumar, S. & Schaal, S. 2001 Learning inverse kinematics. In *IEEE International Conference on Intelligent Robots and Systems (IROS 2001)*. Maui, Hawaii, Oct.29-Nov.3: Piscataway, NJ: IEEE.

Dautenhahn, K. 1995 Getting to know each other—artificial social intelligence for autonomous robots. *Robotics and Autonomous Systems* **16**, 333-356.

Dautenhahn, K. & Nehaniv, C. L. (ed.) 2002 *Imitation in animals and artifacts*. Cambridge, MA: MIT Press.

Dayan, P., Hinton, G. E., Neal, R. M. & Zemel, R. S. 1995 The Helmholtz machine. *Neural Computation* **7**, 889-904.

Demiris, J. & Hayes, G. 2002 Imitation as a dual-route process featuring predictive and learning components: A biologically plausible computational model. In *Imitation in animals and artificats* (ed. K. Dautenhahn & C. L. Nehaniv), pp. 327-361. Cambridge, MA: MIT Press.

Deutscher, J., Blake, A. & Reid, I. 2000 Articulated body motion capture by annealed particle filtering. In *IEEE Computer Vision and Pattern Recognition (CVPR 2000)*. Hilton Head Island, SC, June 13-15, 2000: Piscataway, NJ: IEEE.

Dillmann, R., Kaiser, M. & Ude, A. 1995 Acquisition of elementary robot skills from human demonstration. In *International Symposium on Intelligent Robotic Systems (SIRS'95)*, pp. 1-38. Pisa, Italy, July 10-14, 1995.

Dufay, B. & Latombe, J. C. 1984 An approach to automatic robot programming based on inductive learning. *International Journal of Robotics Research* **3**, 3-20.

Dyer, P. & McReynolds, S. R. 1970 *The computation and theory of optimal control*. New York: Academic Press.

Fagg, A. H. & Arbib, M. A. 1998 Modeling parietal-premotor Interactions in Primate Control of Grasping. *Neural Networks* **11**, 1277-1303.

Gallese, V. & Goldman, A. 1998 Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* **2**, 493-501.

Grudic, G. Z. & Lawrence, P. D. 1996 Human-to-robot skill transfer using the SPORE approximation. In *International Conference on Robotics and Automation*, pp. 2962-2967. Minneapolis, MN, April, 1996: Piscataway, NJ: IEEE.

Guenther, F. H. & Barreca, D. M. 1997 Neural models for flexible control of redundant systems. In *Self-organization, Computational Maps, and Motor Control* (ed. P. Morasso & V. Sanguineti), pp. 102-108. Amsterdam: Elsevier.

Hatsopoulos, N. G. 1996 Coupling the neural and physical dynamics in rhythmic movements. *Neural Comput* **8**, 567-81.

Hayes, G. & Demiris, J. 1994 A robot controller using learning by imitation. In *Procedings of the 2nd International Symposium on Intelligent Robotic Systems* (ed. A. Borkowski & J. L. Crowley), pp. 198-204. Grenoble, France, July 1994: LIFTA-IMAG.

Hovland, G. E., Sikka, P. & McCarragher, B. J. 1996 Skill acquisition from human demonstration using a hidden Markov Model. In *IEEE International Conference on Robotics and Automation*, pp. 2706-2711. Minneapolis, MN, April, 1996: Piscataway, NJ: IEEE.

Ijspeert, J. A., Nakanishi, J. & Schaal, S. 2002a Learning rhythmic movements by demonstration using nonlinear oscillators. In *IEEE International Conference on Intelligent Robots and Systems (IROS 2002)*, pp. 958-963. Lausanne, Sept.30-Oct.4 2002: Piscataway, NJ: IEEE.

Ijspeert, J. A., Nakanishi, J. & Schaal, S. 2002b Movement imitation with nonlinear dynamical systems in humanoid robots. In *International Conference on Robotics and Automation (ICRA2002)*. Washinton, May 11-15 2002.

Ikeuchi, K., Kawade, M. & Suehiro, T. 1993 Assembly task recognition with planar, curved and mechanical contacts. In *Proc. IEEE International Conference on Robotics and Automation,*, vol. 2, pp. 688-693. Atlanta, May 1993: Piscataway, NJ: IEEE.

Kawamura, S. & Fukao, N. 1994 Interpolation for input torque patterns obtained through learning control. In *International Conference on Automation, Robotics and Computer Vision (ICARCV'94)*, pp. 183-191. Singapore, Nov., 1994.

Kawato, M. 1996 Bi-directional theory approach to integration. In *Attention and Performance XVI* (ed. J. Konczak & E. Thelen), pp. 335-367. Cambridge, MA: MIT Press.

Kawato, M. 1999 Internal models for motor control and trajectory planning. *Curr Opin Neurobiol* **9**, 718-727.

Kuniyoshi, Y., Inaba, M. & Inoue, H. 1989 Teaching by showing: Generating robot programs by visual observation of human performance. In *Proceedings of the International Symposium of Industrial Robots*, pp. 119-126. Tokyo, Japan, Oct.4-6.

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**, 799-822.

Lacquaniti, F. & Soechting, J. F. 1986 Simulation studies on the control of posture and movement in a multi-jointed limb. *Biol Cybern* **54**, 367-78.

Levas, A. & Selfridge, M. 1984 A user-friendly high-level robot teaching system. In *IEEE International Conference on Robotics*, pp. 413-416. Altanta, GA, March 1984: Piscataway, NJ: IEEE.

Lin, L.-J. 1991 Programming robots using reinforcement learning and teaching. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, vol. 2, pp. 781-786. Anaheim, CA, July 14-19: Menlo Park, CA: AAAI.

Lozano-Pérez, T. 1982 Task-Planning. In *Robot motion: Planning and control* (ed. M. Brady, J. M. Hollerbach, T. L. Johnson, T. Lozano-Pérez & M. T. Mason), pp. 473-498. Cambridge, MA: MIT Press.

Mataric, M. J. & Pomplun, M. 1998 Fixation behavior in observation and imitation of human movement. *Cogn Brain Res* **7**, 191-202.

Meltzoff, A. N. & Moore, M. K. 1994 Imitation, memory, and the represesntation of persons. *Infant Behavior and Development* **17**, 83-99.

Meltzoff, A. N. & Moore, M. K. 1997 Explaining facial imitation: A theoretical model. *Early Development and Parenting* **6**, 179-192.

Miall, R. C. & Wolpert, D. M. 1996 Forward models for physiological motor control. *Neural Networks* **9**, 1265-1285.

Miyamoto, H. & Kawato, M. 1998 A tennis serve and upswing learning robot based on bi-directional theory. *Neural Networks* **11**, 1331-1344.

Miyamoto, H., Schaal, S., Gandolfo, F., Koike, Y., Osu, R., Nakano, E., Wada, Y. & Kawato, M. 1996 A Kendama learning robot based on bi-directional theory. *Neural Networks* **9**, 1281-1302.

Nechyba, M. C. & Xu, Y. 1995 Human skill transfer: neural networks as learners and teachers. In *IEEE/RSJ International Conference on Intelligence Robots and Systems*, vol. 3, pp. 314-319. Pittsburgh, PA, August 5-9: Piscataway, NJ: IEEE.

Nehaniv, C. L. & Dautenhahn, K. 1999 Of hummingbirds and helicopters: An algebraic framework for interdisciplinary studies of imitation and its applications. In *Learning robots: An interdisciplinary approach* (ed. J. Demiris & A. Birk): World Scientific.

Oztop, E. & Arbib, M. A. 2002 Schema design and implementation of the grasp-related mirror neuron system. *Biol Cybern* **87**, 116-40.

Pawelzik, K., Kohlmorgen, J. & Müller, K. R. 1996 Annealed competition of experts for a segmentation and classification of switching dynamics. *Neural Computation* **8**, 340-356.

Piaget, J. 1951 *Play, dreams, and imitation in childhood*. New York,: Norton.

Pook, P. K. & Ballard, D. H. 1993 Recognizing teleoperated manipulations. In *Proceedings IEEE International Conference on Robotics and Automation*, vol. 3, pp. 913-918. Atlanta, GA, May 1993: Piscataway, NJ: IEEE.

Rittscher, J. & Blake, A. submitted Mathematical modelling of animate and intentional motion. *Philosophical Transactions of the Royal Society of London B*.

Rizzolatti, G. & Arbib, M. A. 1998 Language within our grasp. *Trends Neurosci* **21**, 188-94.

Rizzolatti, G., Fadiga, L., Gallese, V. & Fogassi, L. 1996 Premotor cortex and the recognition of motor actions. *Cognitive Brain Research* **3**, 131-141.

Saltzman, E. & Kelso, S. J. A. 1987 Skilled actions: A task-dynamic approach. *Psychological Review* **94**, 84-106.

Sammut, C., Hurst, S., Kedzier, D. & Michie, D. 1992 Learning to fly. In *Proceedings of the Ninth International Machine Learning Conference (ML'92)* (ed. D. Sleeman & P. Edwards), pp. 385-393. Aberdeen, Scotland, July 1-3: San Mateo, CA: Morgan Kaufman.

Schaal, S. 1997 Learning from demonstration. In *Advances in Neural Information Processing Systems 9* (ed. M. C. Mozer, M. Jordan & T. Petsche), pp. 1040-1046. Cambridge, MA: MIT Press.

Schaal, S. 1999 Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences* **3**, 233-242.

Schaal, S., Sternad, D., Dean, W., Kotoska, S., Osu, R. & Kawato, M. 2000 Reciprocal excitation between biological and robotic research. In *Sensor Fusion and Decentralized Control in Robotic Systems III*, *Proceedings of SPIE*, vol. 4196, pp. 30-40. Boston, MA, Nov.5-8, 1992.

Schwenker, F., Sommer, F. T. & Palm, G. 1996 Iterative retrieval of sparsely coded associateve memory patterns. *Neural Networks* **9**, 445-455.

Segre, A. B. & DeJong, G. 1985 Explanation-based manipulator learning: Acquisition of planning ability through observation. In *IEEE Conference on Robotics and Automation*, pp. 555-560. St. Louis, MO, March 1985: Piscataway, NJ: IEEE.

Segre, A. M. 1988 *Machine learning of robot assembly plans*. Kluwer international series in engineering and computer science. Knowledge representation, learning, and expert systems. Boston: Kluwer Academic Publishers.

Sternad, D. & Schaal, D. 1999 Segmentation of endpoint trajectories does not imply segmented control. *Experimental Brain Research* **124**, 118-136.

Sutton, R. S. & Barto, A. G. 1998 *Reinforcement learning : An introduction.* Adaptive computation and machine learning. Cambridge: MIT Press.

Sutton, R. S., Singh, S., Precup, D. & Ravindran, B. 1999 Improved switching among temporally abstract actions. In *Advances in Neural Information Processing Systems 11*: MIT Press.

Tomasello, M., Savage-Rumbaugh, S. & Kruger, A. C. 1993 Imitative learning of actions on objects by children, chimpanzees, and enculturated chimpanzees. *Child Dev* **64**, 1688-705.

Tung, C. P. & Kak, A. C. 1995 Automatic learning of assembly task using a DataGlove system. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1, pp. 1-8. Pittsburgh, PA.

van Essen, D. C. & Maunsell, J. M. R. 1983 Hierachical organization and functional streams in the visual cortex. *Trends in Neuroscience* **6**, 370-375.

Wada, Y. & Kawato, M. 1995 A theory for cursive handwriting based on the minimization principle. *Biological Cybernetics* **73**, 3-13.

Weigend, A. S., Mangeas, M. & Srivastava, A. N. 1995 Nonlinear gated experts for time series: discovering regimes and avoiding overfitting. *International Journal of Neural Systems* **6**, 373-399.

Widrow, B. & Smith, F. W. 1964 Pattern recognizing control systems. In *1963 Computer and Information Sciences (COINS) Symp. Proc.*, pp. 288-317. Spartan.

Wolpert, D. M. 1997 Computational approaches to motor control. *Trends in Cognitive Sciences* **1**, 209-216.

Wolpert, D. M., Doya, K. & Kawato, M. submitted A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London B*.

Wolpert, D. M. & Kawato, M. 1998 Multiple paired forward and inverse models for motor control. *Neural Networks* **11**, 1317-1329.

Wolpert, D. M., Miall, R. C. & Kawato, M. 1998 Internal models in the cerebellum. *Trends in Cognitive Sciences* **2**, 338-347.