A Brief Overview of Human and Robot Motor Learning

Ruta Desai
rutad@cs.cmu.edu

Motor learning and adaptation has interested researchers from Neuroscience, Robotics and Statistics over decades. Motor learning involves complex interplay of memory, plasticity, perception and muscle control. The focus of current review is on algorithms that could be used to model learning of control in motor tasks especially in locomotion domain.

1 Human Motor Learning

Motor learning in humans has been studied mostly through adaptation in motor tasks (by introducing perturbation in a task). Motor adaptation tasks include arm reaching movements, finger grasps and aiming tasks [1]. Perturbation are introduced in reaching or aiming movements using either force fields or visuomotor transformations. Various computational approaches have been proposed to model control and adaptation in these motor tasks.

1.1 Models for Motor Adaptation

Three prominent models in the literature are *Optimal models*, *Bayesian models* and *Motor Primitive* based models [2, 3, 4]. Optimal models assume that humans are optimal with respect to a cost function. Todorov and colleagues [5, 6] propose the minimum intervention principle, suggesting that motor control reduces variability of motion only in direction of goal for goal directed reaching movements using optimal control. Other researchers have proposed minimization of effort, jerk and motor noise to explain human motion [7, 8]. Efforts have been also made to understand hypothetical cost functions that humans might be optimizing for these motor tasks by comparing a number of cost functions with human data [9]. Motor learning and adaptation in optimal control framework can thus be looked upon as acquisition of features and cost function and deriving the optimal policy thereafter. Bayesian models have been mostly used to explain sensorimotor integration. Kording and Wolpert [10] hypothesize that humans deal with uncertainty in perception and action in a Bayesian manner. They found that human subjects learn prior distribution of the discrepancy in a reaching task and estimate the reliability of visual information (likelihood). The subjects combined these two sources of information in a Bayesian manner. In Bayesian framework, motor learning then comprises of acquisition of priors. Bayesian models easily allow combining prior knowledge with evidence. *Motor primitive* based models have been quite popular as well. Sing et. al. [11] propose a visco-elastic primitive model where the basis functions are composed of position and velocity of the end-effector. These functions combine in weighted proportion to produce the desired force needed against force-field perturbation in a reaching task. Thoroughman and Shadmehr [12] propose gaussian like functions of hand velocity as basis functions encoding an internal model transforming desired arm trajectories into muscle forces to explain adaptation in force-fields. Weights for these primitives are learnt by first or second order gradient methods [11, 12] . At the level of muscle forces this approach resembles the concept of muscle synergies (pattern of co-activation of muscles using a single neural command). Bizzi et. al. [13] find muscle synergies to explain Electro-myograph (EMG) signals in many tasks like jumping and walking in frogs although they do not extend this to adaptation or learning. Recently, Dominici, Ivenenko and colleagues [14], extended the idea of muscle synergies to explain motor learning. The authors extract muscle synergy patterns from neonates, toddlers, pre-schoolers and adults during walking. They find that sinusoidal synergies could explain EMG signals in neonates and toddlers. They also find a transitional change in synergies from toddlers to preschooler and finally to adults, hinting towards continuous development of motor modules. Interestingly, similarities in synergies were also found in human, monkey, cat and rat toddlers, suggesting that primitives are conserved over morphological differences and phylogenetic distances. This work seems most relevant to our interest of modeling stages of learning although the authors do not explain the mechanism for transformation of synergies.

1.2 Goals of Motor Learning

Another way of looking at motor learning literature is to ask what is it that the humans are learning. This is based on the fact that motor control requires dealing with some inherent characteristics of human motion like uncertainty, noise and delays. A major school of thought that was developed to explain human motion in presence of noise and motor delays uses internal models: forward prediction models (which predict the desired motion) and inverse control models (which map predicted desired motion into corresponding control inputs). In other words these models encode *forward and inverse dynamics*. Evidence of such models have been found in anticipatory postural adjustment based on known arm dynamics during lifting and anticipatory grip force adjustment in finger grasps [15, 16]. Motor learning is then modeled as acquisition and update of these models. Haruno and colleagues propose one such model called MOSAIC model for motor learning and control where they learn forward and inverse models using gradient based or Expectation-Maximization techniques [17]. Researchers have also tried to hypothesize about generalization of control in similar tasks using internal models. Berniker and Kording suggest that nervous system might be performing credit assignment to different sources of errors (body or environment) to figure out how to generalize [18]. For example, estimated changes of limb properties will affect movements across the workspace but not movements of the other limb. They solve the credit assignment by Bayesian inference using internal models of limb dynamics and the environment in force-field experiments. Some researchers prefer to think of internal model as an ideal *feedforward controller* for a given task instead of task dynamics [19]. Many times such feedforward controllers are learnt by feedback-error-learning or auto-regressive models using trial-by-trial error [20]. Neural networks have also been used for representation and learning of such controllers [21]. Along with dynamics and feedforward control, humans can also learn *kinematic transformations* between joint space and task space. This was shown by Mussa-Ivaldi and colleagues in experiments involving target aiming on screen where the cursor was controlled by linearly mapped joint motion of the hands [22, 23]. To deal with uncertainty of the environment it has been proposed that humans learn to control *impedance* of the endpoint of their limb [24]. Iterative learning can be used to learn such impedances [25]. Impedance control can not only help during instability and uncertainty of the environment but can also help when there is uncertainty in the internal model. (During learning of novel dynamics, initial stiffness is higher to maintain stability. This stiffness reduces as the dynamics model gets better.) This suggests that some of the above models might be working in parallel [26]. Number of papers later extended the idea of impedance control to account for metabolic cost, task-relevant variability along with stability [27, 28].

1.3 Mechanisms for Motor Adaptation

Researchers have also tried to explain the various mechanisms involved in motor learning. Three main mechanisms found in the literature are: Error based learning, reinforcement learning and Structural learning [1]. Error based learning uses an error signal to guide the learning process. Gradient based updates, feedback error learning are some examples of error based learning (used in many of the approaches listed above). Apart from these traditional techniques, Franklin and colleagues proposed a V-shaped function to update the feedforward muscle command by using trial-by trial errors in reaching movements reflected by change in muscle lengths [29]. They could explain EMG patterns of arm muscles during reaching-movements in force fields using their proposed V-shaped learning function. Reinforcement learning uses a relative measure of success or failure as a guiding signal for learning. Such an unsigned signal will not have a gradient/directional information. However, in situations in which a complex sequence of actions need to take place to achieve a goal and the outcome or reward is not immediately reflected from the action (for example, learning the movements required to make a playground swing go higher), error-based learning cannot easily be applied and reinforcement learning techniques can be used to assign credit or blame, back in time, to the actions that led to success or failure. Izawa and Shadmehr [30] show that reinforcement learning can explain adaptation by using sensory and reward prediction scalar errors to drive motor adaptation using Temporal Difference (TD) method. They compared learning between two groups one which received sensory feedback with explicit reward and one which received only explicit scalar reward in a visuomotor rotation task. They find that both groups learn comparably except that the group which received sensory feedback showed generalization across directions. They hypothesize that scalar reward makes the subjects use a trial-and-error exploration strategy for learning. A combination of model based learning using errors and model free learning using reinforcement has also been suggested for adaptation in visuomotor rotation task [31]. In Structural learning, a task is represented by three levels: the structure of the task, its parameters and the relevant state. The structure represents the relevant inputs and outputs of the system and the functional form of the equations that relate them. For example, when we learn to play tennis we have to identify the task-relevant inputs and outputs – such as arm motor commands and racquet head motion – as well as the mapping between them, which depends on the geometry and dynamics of the racquet. Learning the structure in one task, such as tennis, can be beneficial for tasks that share a similar structure, such as other racquet sports. What differs between tennis, squash and badminton are the parameters of the structures, such as the racquet length, head size and weight. In principle, if the structure is known then the parameters of the system can be quickly identified, allowing rapid learning [32]. Braun and colleagues [32] pointed out the important distinction between parametric learning and structural learning. Parametric learning describes the adaptation processes: countering a perturbation through error-driven updates of a parameterized model. Structural learning can be considered learning the covariance structure among these parameters. In support of their hypothesis, they show that new structures can be learnt by exposing participants to a randomly varying set of tasks that share a common structure but vary in their parameter settings. For example, after being exposed to horizontal or vertical visuomotor rotations in a three-dimensional reaching task, participants adapted more rapidly to new tasks that share the same structure [33]. Structure can also be modeled as a Bayesian network or topology of dependent variables and then 'structural learning' would refer to learning the topology of such a network, whereas 'parametric learning' could be employed to determine quantitatively the causal connections given by the structure [32]. Temporal variations in structure could be modeled using Hidden Markov Models [32]. The role of use-dependent plasticity (repetition based learning) and explicit cognitive processes has also been explored by some researchers. Dierdrichsen and colleagues [34] show that even in error based tasks like reaching in force-fields, use-dependent plasticity can occur along with error-based adaptation. They model use-dependent plasticity as a simple averaging process between last planned and last actual movement of hand. Gentner and colleagues [35] perform Principal Component Analysis (PCA) to obtain joint correlation patterns in finger movements evoked by transcranial magnetic stimulation over the primary motor cortex. Linear combinations of a selected subset of these patterns were used to reconstruct active instrumental playing or grasping movements. They find that reconstruction quality of instrumental playing was superior in skilled musicians compared to musically untrained subjects, indicating that experience-dependent motor skills are specifically encoded in the functional organization of the primary motor cortex.

1.4 Motor Skill Learning vs Motor Adaptation

Skill learning is quite different from adaptation because unlike adaptation baseline behavior is yet to be acquired in skill learning. While adaptation can be measured as reduction of error, such a metric can not be used for skill learning because of its highly unstructured nature. Some researchers have tried to use speed-accuracy tradeoff function and reduction in variability of movement to define skill learning [36]. However, mechanisms under play for skill learning are speculative. In a review of various mechanisms of motor learning and adaptation, Krakauer and Mazzoni [37] argue that it is hard to say if any of the mechanisms for adaptation could be applied for skill learning. Authors suggest that motor learning, at the very least, is made up of adaptation, use-dependent plasticity, operant reinforcement, and explicit cognitive processes (all the mechanisms stated above working together). In this framework, they also conjecture that adaptation and skill learning tasks lie along a spectrum with model-based processes prominent in the former and model-free processes prominent in the latter.The authors are of opinion that in model-based approaches, Optimal Feedback Control [2] comes closest in explaining some of the above mentioned metrics of skill learning. However, most times such frameworks succeed in establishing correlation of simulated data with the observed data but fail at prediction for other similar tasks.

1.5 Outlook

Finally, there are two other prominent issues with the human motor learning literature. Firstly most of the test tasks are related to structured arm movements and are not highly dynamic such as tasks in locomotion domain (walking, skating and so on). Secondly, algorithms and mechanisms suggested for motor learning do not output feedback structures or policies in a conventional sense of control theory. In other words, no research addresses learning or adaptation of muscle feedback control policies (muscle reflexes). However, this seems to have been of prime interest in robotics community for robot learning.

2 Robot Motor Learning

2.1 Overview

Schaal and Atkeson [38] provide a good classification of learning control in Robotics. They divide the field based on methods used, type of tasks and direct (model-free) vs indirect (model-based) control. Methods in which a model is first identified before learning of control are called ‘model-based’ learning methods. While methods where the policy is learnt directly without the detour into model identification are called ‘model-free’ learning methods.

2.2 Imitation Learning

A big class of motor learning in robotics is learning from demonstration or imitation learning with the idea of using demonstration data to make the robot perform a task [39]. Two main entities that are learnt from demonstration data are: mapping function between states to action space (policy/controller) and system model (dynamics and reward/cost functions).

2.2.1 Learning Policies

Because of its nature many supervised learning techniques find use in learning the mapping function using demonstration data. Pomerleau trained a neural network using human demonstrations to drive an autonomous van on variety of roads [40]. Neural networks have also been used to enable a robot arm to do the peg in the hole task [41]. In a prominent work, Billard and Mataric [42] trained hierarchical neural networks inspired by biological architecture to imitate human arm reaching movements with a biomechanical model. Decision trees have been used in an inductive program to learn to fly in a flight simulator using human data [43]. Number of regression techniques like Locally Weighted Regression(LWR) [44], Receptive Field Weighted Regression (RCWR) [45], Locally Weighted Projection Regression (LWPR) [46] have been used for learning the mapping function. In essence, these techniques find, in the spirit of a first-order Taylor series expansion, the linearization of the function at an input point, and the region (also called a kernel) in which this linearization holds within a certain error bound.

Instead of learning mapping functions (policies) from scratch, Ijspeert and colleagues [47] encode inherent nature of a policy by representing policies using differential equations with well-defined non-linear attractor properties, called dynamic motion primitives (DMP). Without learning the model, DMPs allow generalization and self-improvement [48, 49, 50]. DMPs can model control policies for both discrete and rhythmic motions and have been successfully used to learn tasks such as ball in the cup, ball paddling, playing tennis, drumming and so on from demonstrations [51]. Another representation that has been used in robot learning community for policies is motor primitives or smaller units of motion that combine together for motion skills. Bentivegna and colleagues [52] manually build a library of basic motions to be used as primitives and then use a classifier trained from demonstration to select a primitive for air hockey and tilt maze task. Using similar approach, Nakaoaka and colleagues imitate human dancers by decomposing the motion capture data into manually defined task primitives with appropriate parameters to maintain dynamic feasibility [53]. Instead of manually defining the primitives, Jenkins and Mataric [54] define primitives as spatio-temporal clusters and extract them from motion capture. They also extract meta-level behaviors to link the primitives from the data for humanoid motion synthesis.

One issue with such direct control methods is that the task goal is hidden and hence the learnt mapping function cannot be used for generalization or improvement (exception being policies learnt as DMPs). Improvement might be essential because the demonstration data would cover only limited state space. Model-based methods or indirect methods are one way of dealing with this (Learn the model of task instead of the mapping function directly.)

2.2.2 Learning Models

Atkeson and Schaal [55] used parametric and nonparametric regression using some of the techniques mentioned above for learning a model of dynamics and then used an optimal controller (Linear Quadratic Regulator (LQR)) to make a robot balance an inverted pole. Nonlinear approximation methods have also been proposed using Gaussian processes [56, 57]. Some non-linear regression techniques have been tried in reduced dimensions, where regression maps the high dimensional data to reduced dimensions and back and a predictive model may be learnt in the reduced space [58, 59]. For non-linear models, one could use conventional dynamic programming approaches to obtain a controller [38]. While Dynamic programming is a global optimizer, it might not be practical to do Dynamic programming because of the curse of dimensionality. A locally optimal equivalent is called Trajectory libraries, where locally optimal sequence of commands (policy) is learnt from various initial conditions using optimizers like SNOPT [60]. Learning the model of dynamics can also be compared to some approaches in human motor learning which hypothesize about learning through internal models (see section: 1). While Atkeson and Schaal [55] learn linearized forward dynamic model, researchers have also tried to learn inverse models because of the popularity of operational space control in robotics [61]. Because such inverse models can be one-to-many relations, Peters and Schaal [61] formulate it as reinforcement learning reward optimization problem and use Expectation-Maximization to solve it. Paired forward-inverse model is constantly adapted using regression to do pendulum swing up task by Vaandrager and colleagues [62]. This is similar to MOSAIC model (see section 1) except that a single model is adapted instead of using multiple paired models.

2.2.3 Learning Rewards

Apart from policies and dynamic models, rewards can be learnt from demonstrations. Rewards can help one to infer the intent of the expert or the features that the expert cares about. Inverse Reinforcement Learning (IRL) or Inverse Optimal Control (IOC) techniques have been used for this [63, 64]. Kalakrishnan and colleagues [65] propose a path integral approach to learn convex objective function for manipulation. Other similar approaches include Maximum and Relative Entropy IRL [66, 67]. All these methods use Markov Decision Processes (MDP) and also try to deal with issues in imitation learning like noise in data, limited state space exploration in the demonstrated data and so on. These approaches have shown successful learning for tasks like driving a race car in simulation and ball-in-the-cup task. While these methods can infer a cost/reward function from demonstrated data assuming optimal behavior of the expert/demonstration, there is no guarantee that it is the true function which the expert was using. Further, most cost functions that are learnt are either linear combination of features or convex functions, for ease of computation.

2.2.4 Statistical Models of Motor Skills

Instead of learning policies, models or rewards, some researchers have focussed on statistical modeling of skill for a task. Hovland and colleagues discretize a demonstrated skill into important events modeled by a Hidden Markov Model (HMM) with each state of HMM providing reference command for an underlying robot controller in a planar peg insertion task [68]. HMMs have been very popular for skill encoding because of its success with time-series data modeling in domains like speech. Lee and Nakamura [69] use HMM as a primitive and also couple it with a likelihood indicating a primitive's contribution in generating a motion pattern. HMM have been used together with Gaussian Mixture Model (GMM) to achieve a spatio-temporal encoding of skill which can be reconstructed using Gaussian Mixture Regression [70]. Authors also present an overall framework for imitation in terms of: what to imitate (probabilistic encoding with HMM and GMM), metric of imitation and how to imitate (optimal trajectory generation) and demonstrate it in a pick and place task. In another work, the same authors have also used PCA and HMM together [71]. Such approaches have also shown some success with generalization of skills. A good overview of skill encoding as symbolic, statistical models and dynamical systems can be found in [72]. Statistical approaches mostly model only kinematic spatio-temporal variations, neglecting dynamics which might not be ideal for many motor tasks [70]. Moreover, such methods also suffer from the curse of dimensionality. Robustness of such approaches in presence of disturbances is also unclear.

Some researchers have felt that expert demonstrations might not contain necessary information for learning a skill and hence they looked at failed demonstrations. They developed probabilistic approaches (using mixture models) that avoid reproducing observed failures while leveraging the variance across multiple attempts to drive exploration in task which involved throwing ball in a basket [73].

There are number of works which talk about generalization of motor skills as well as transfer of skills but they are not reviewed here.

2.3 Reinforcement Learning

Along with imitation learning, another prominent approach used for robot learning is through reinforcement. Most of reinforcement learning methods are model-free. Value function or action value function called Q-function can be learnt with Temporal Difference learning [74]. Once the value function or Q-function is learnt, policy can be easily obtained.

Again, to deal with the curse of dimensionality, approximation methods such as fitted Q-learning have been proposed [75]. Some approaches do not estimate the value function but rather focus directly on learning the control policy from trajectory rollouts such as policy gradient methods [76]. Parameterized policies have been learnt using rollouts on hardware as well [77]. Gradient free black box optimization approaches have also achieved good success in learning complex motion skills. Liu and colleagues learn various bicycle stunts by learning feedforward controllers parameterized as splines and feedback controllers represented by neural network using black-box policy search [78]. Linear feedback policies have also been learnt in similar fashion for a variety of other tasks such as balancing on a board, walking, running and juggling [79].

2.4 Outlook

While a large body of work in robot learning talks about imitating human motion kinematically and dynamically, very few approaches relate to biological findings and hypotheses about human motor control. Further more, most imitation learning research use human data as expert data and employs exploration or self-improvement later if needed. No research has truly aimed at modeling stages of learning a motor skill in humans (modeling humans as naive learners rather than experts). Ideally we wish that the use of such algorithms will help us in making predictions about human motor learning. Some of these methods have been used for modeling human motor adaptation (see section: 1 for more details) but not for complex motor skills.

References

[1] D. M. Wolpert, J. Diedrichsen, and J. R. Flanagan, "Principles of sensorimotor learning," *Nature Reviews Neuroscience*, vol. 12, no. 12, pp. 739–751, 2011.

[2] E. Todorov, "Optimality principles in sensorimotor control," *Nature neuroscience*, vol. 7, no. 9, pp. 907–915, 2004.

[3] K. P. Körding and D. M. Wolpert, "Bayesian decision theory in sensorimotor control," *Trends in cognitive sciences*, vol. 10, no. 7, pp. 319–326, 2006.

[4] C. Alessandro, I. Delis, F. Nori, S. Panzeri, and B. Berret, "Muscle synergies in neuroscience and robotics: from input-space to task-space perspectives," *Frontiers in computational neuroscience*, vol. 7, 2013.

[5] E. Todorov and M. I. Jordan, "Optimal feedback control as a theory of motor coordination," *Nature neuroscience*, vol. 5, no. 11, pp. 1226–1235, 2002.

[6] D. Liu and E. Todorov, "Evidence for the flexible sensorimotor strategies predicted by optimal feedback control," *The Journal of Neuroscience*, vol. 27, no. 35, pp. 9354–9368, 2007.

[7] M. Haruno and D. M. Wolpert, "Optimal control of redundant muscles in step-tracking wrist movements," *Journal of Neurophysiology*, vol. 94, no. 6, pp. 4244–4255, 2005.

[8] T. Flash and N. Hogan, "The coordination of arm movements: an experimentally confirmed mathematical model," *The journal of Neuroscience*, vol. 5, no. 7, pp. 1688–1703, 1985.

[9] M. Berniker, M. K. O´ Brien, K. P. Kording, and A. A. Ahmed, "An examination of the generalizability of motor costs," *PloS one*, vol. 8, no. 1, p. e53759, 2013.

[10] K. P. Körding and D. M. Wolpert, "Bayesian integration in sensorimotor learning," *Nature*, vol. 427, no. 6971, pp. 244–247, 2004.

[11] G. C. Sing, W. M. Joiner, T. Nanayakkara, J. B. Brayanov, and M. A. Smith, "Primitives for motor adaptation reflect correlated neural tuning to position and velocity," *Neuron*, vol. 64, no. 4, pp. 575–589, 2009.

[12] K. A. Thoroughman and R. Shadmehr, "Learning of action through adaptive combination of motor primitives," *Nature*, vol. 407, no. 6805, pp. 742–747, 2000.

[13] E. Bizzi, V. Cheung, A. d'Avella, P. Saltiel, and M. Tresch, "Combining modules for movement," *Brain Research Reviews*, vol. 57, no. 1, pp. 125–133, 2008.

[14] N. Dominici, Y. P. Ivanenko, G. Cappellini, A. d?Avella, V. Mondì, M. Cicchese, A. Fabiano, T. Silei, A. Di Paolo, C. Giannini *et al.*, "Locomotor primitives in newborn babies and their development," *Science*, vol. 334, no. 6058, pp. 997–999, 2011.

[15] P. R. Davidson and D. M. Wolpert, "Widespread access to predictive models in the motor system: a short review," *Journal of Neural Engineering*, vol. 2, no. 3, p. S313, 2005.

[16] J. R. Flanagan, P. Vetter, R. S. Johansson, and D. M. Wolpert, "Prediction precedes control in motor learning," *Current Biology*, vol. 13, no. 2, pp. 146–150, 2003.

[17] M. Haruno, D. Wolpert, and M. Kawato, "Mosaic model for sensorimotor learning and control," *Neural computation*, vol. 13, no. 10, pp. 2201–2220, 2001.

[18] M. Berniker and K. Kording, "Estimating the sources of motor errors for adaptation and generalization," *Nature neuroscience*, vol. 11, no. 12, pp. 1454–1461, 2008.

[19] R. Shadmehr and F. A. Mussa-Ivaldi, "Adaptive representation of dynamics during learning of a motor task," *The Journal of Neuroscience*, vol. 14, no. 5, pp. 3208–3224, 1994.

[20] R. A. Scheidt, J. B. Dingwell, and F. A. Mussa-Ivaldi, "Learning to move amid uncertainty," *Journal of neurophysiology*, vol. 86, no. 2, pp. 971–985, 2001.

[21] M. Kawato, K. Furukawa, and R. Suzuki, "A hierarchical neural-network model for control and learning of voluntary movement," *Biological cybernetics*, vol. 57, no. 3, pp. 169–185, 1987.

[22] K. M. Mosier, R. A. Scheidt, S. Acosta, and F. A. Mussa-Ivaldi, "Remapping hand movements in a novel geometrical environment," *Journal of neurophysiology*, vol. 94, no. 6, pp. 4362–4372, 2005.

[23] X. Liu, K. M. Mosier, F. A. Mussa-Ivaldi, M. Casadio, and R. A. Scheidt, "Reorganization of finger coordination patterns during adaptation to rotation and scaling of a newly learned sensorimotor transformation," *Journal of neurophysiology*, vol. 105, no. 1, pp. 454–473, 2011.

[24] E. Burdet, R. Osu, D. W. Franklin, T. E. Milner, and M. Kawato, "The central nervous system stabilizes unstable dynamics by learning optimal impedance," *Nature*, vol. 414, no. 6862, pp. 446–449, 2001.

[25] E. Burdet, K. P. Tee, I. Mareels, T. E. Milner, C.-M. Chew, D. W. Franklin, R. Osu, and M. Kawato, "Stability and motor adaptation in human arm movements," *Biological cybernetics*, vol. 94, no. 1, pp. 20–32, 2006.

[26] D. W. Franklin, R. Osu, E. Burdet, M. Kawato, and T. E. Milner, "Adaptation to stable and unstable dynamics achieved by combined impedance control and inverse dynamics model," *Journal of neurophysiology*, vol. 90, no. 5, pp. 3270–3282, 2003.

[27] D. W. Franklin and D. M. Wolpert, "Specificity of reflex adaptation for task-relevant variability," *The Journal of Neuroscience*, vol. 28, no. 52, pp. 14 165–14 175, 2008.

[28] D. W. Franklin, U. So, M. Kawato, and T. E. Milner, "Impedance control balances stability with metabolically costly muscle activation," *Journal of neurophysiology*, vol. 92, no. 5, pp. 3097–3105, 2004.

[29] D. W. Franklin, E. Burdet, K. P. Tee, R. Osu, C.-M. Chew, T. E. Milner, and M. Kawato, "Cns learns stable, accurate, and efficient movements using a simple algorithm," *The Journal of Neuroscience*, vol. 28, no. 44, pp. 11 165–11 173, 2008.

[30] J. Izawa and R. Shadmehr, "Learning from sensory and reward prediction errors during motor adaptation," *PLoS computational biology*, vol. 7, no. 3, p. e1002012, 2011.

[31] V. S. Huang, A. Haith, P. Mazzoni, and J. W. Krakauer, "Rethinking motor learning and savings in adaptation paradigms: model-free memory for successful actions combines with internal models," *Neuron*, vol. 70, no. 4, pp. 787–801, 2011.

[32] D. A. Braun, C. Mehring, and D. M. Wolpert, "Structure learning in action," *Behavioural brain research*, vol. 206, no. 2, pp. 157–165, 2010.

[33] D. A. Braun, A. Aertsen, D. M. Wolpert, and C. Mehring, "Motor task variation induces structural learning," *Current Biology*, vol. 19, no. 4, pp. 352–357, 2009.

[34] J. Diedrichsen, O. White, D. Newman, and N. Lally, "Use-dependent and error-based learning of motor behaviors," *The Journal of neuroscience*, vol. 30, no. 15, pp. 5159–5166, 2010.

[35] R. Gentner, S. Gorges, D. Weise, M. Buttmann, J. Classen *et al.*, "Encoding of motor skill in the corticomuscular system of musicians," *Current Biology*, vol. 20, no. 20, pp. 1869–1874, 2010.

[36] H. Müller and D. Sternad, "Decomposition of variability in the execution of goal-oriented tasks: three components of skill improvement." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 30, no. 1, p. 212, 2004.

[37] J. W. Krakauer and P. Mazzoni, "Human sensorimotor learning: adaptation, skill, and beyond," *Current opinion in neurobiology*, vol. 21, no. 4, pp. 636–644, 2011.

[38] S. Schaal and C. G. Atkeson, "Learning control in robotics," *Robotics & Automation Magazine, IEEE*, vol. 17, no. 2, pp. 20–29, 2010.

[39] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and autonomous systems*, vol. 57, no. 5, pp. 469–483, 2009.

[40] D. A. Pomerleau, "Efficient training of artificial neural networks for autonomous navigation," *Neural Computation*, vol. 3, no. 1, pp. 88–97, 1991.

[41] R. Dillmann, M. Kaiser, and A. Ude, "Acquisition of elementary robot skills from human demonstration," in *International Symposium on Intelligent Robotics Systems*, 1995, pp. 185–192.

[42] A. Billard and M. J. Matarić, "Learning human arm movements by imitation:: Evaluation of a biologically inspired connectionist architecture," *Robotics and Autonomous Systems*, vol. 37, no. 2, pp. 145–160, 2001.

[43] C. Sammut, S. Hurst, D. Kedzier, and D. Michie, "Learning to fly," in *In Proceedings of the Ninth International Conference on Machine Learning*, 1992.

[44] C. G. Atkeson, A. W. Moore, and S. Schaal, "Locally weighted learning for control," in *Lazy learning*. Springer, 1997, pp. 75–113.

[45] S. Schaal and C. G. Atkeson, "Constructive incremental learning from only local information," *Neural Computation*, vol. 10, no. 8, pp. 2047–2084, 1998.

[46] S. Vijayakumar and S. Schaal, "Locally weighted projection regression: An o (n) algorithm for incremental real time learning in high dimensional space," in *International conference on machine learning, proceedings of the sixteenth conference*, 2000.

[47] A. J. Ijspeert, J. Nakanishi, and S. Schaal, "Learning attractor landscapes for learning motor primitives," Tech. Rep., 2002.

[48] J. Kober and J. Peters, "Learning motor primitives for robotics," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE, 2009, pp. 2112–2118.

[49] P. Pastor, H. Hoffmann, T. Asfour, and S. Schaal, "Learning and generalization of motor skills by learning from demonstration," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE, 2009, pp. 763–768.

[50] A. Ude, A. Gams, T. Asfour, and J. Morimoto, "Task-specific generalization of discrete and periodic dynamic movement primitives," *Robotics, IEEE Transactions on*, vol. 26, no. 5, pp. 800–815, 2010.

[51] A. J. Ijspeert, J. Nakanishi, H. Hoffmann, P. Pastor, and S. Schaal, "Dynamical movement primitives: learning attractor models for motor behaviors," *Neural computation*, vol. 25, no. 2, pp. 328–373, 2013.

[52] D. C. Bentivegna, C. G. Atkeson, and G. Cheng, "Learning tasks from observation and practice," *Robotics and Autonomous Systems*, vol. 47, no. 2, pp. 163–169, 2004.

[53] S. Nakaoka, A. Nakazawa, F. Kanehiro, K. Kaneko, M. Morisawa, and K. Ikeuchi, "Task model of lower body motion for a biped humanoid robot to imitate human dances," in *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*. IEEE, 2005, pp. 3157–3162.

[54] O. C. Jenkins and M. J. Mataric, "Automated derivation of behavior vocabularies for autonomous humanoid motion," in *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*. ACM, 2003, pp. 225–232.

[55] C. G. Atkeson and S. Schaal, "Robot learning from demonstration," in *ICML*, vol. 97, 1997, pp. 12–20.

[56] D. B. Grimes, D. R. Rashid, and R. P. Rao, "Learning nonparametric models for probabilistic imitation," in *Advances in Neural Information Processing Systems*, 2006, pp. 521–528.

[57] D. Nguyen-Tuong, J. R. Peters, and M. Seeger, "Local gaussian process regression for real time online model learning," in *Advances in Neural Information Processing Systems*, 2009, pp. 1193–1200.

[58] A. P. Shon, K. Grochow, and R. P. Rao, "Robotic imitation from human motion capture using gaussian processes," in *Humanoid Robots, 2005 5th IEEE-RAS International Conference on*. IEEE, 2005, pp. 129–134.

[59] R. Chalodhorn, D. B. Grimes, G. Y. Maganis, R. P. Rao, and M. Asada, "Learning humanoid motion dynamics through sensory-motor mapping in reduced dimensional spaces," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*. IEEE, 2006, pp. 3693–3698.

[60] C. Liu and C. G. Atkeson, "Standing balance control using a trajectory library," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. IEEE, 2009, pp. 3031–3036.

[61] J. Peters and S. Schaal, "Learning to control in operational space," *The International Journal of Robotics Research*, vol. 27, no. 2, pp. 197–212, 2008.

[62] M. Vaandrager, R. Babuska, L. Busoniu, and G. A. Lopes, "Imitation learning with non-parametric regression," in *Automation Quality and Testing Robotics (AQTR), 2012 IEEE International Conference on*. IEEE, 2012, pp. 91–96.

[63] A. Y. Ng, S. J. Russell *et al.*, "Algorithms for inverse reinforcement learning." in *International Conference on Machine learning*, 2000, pp. 663–670.

[64] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 1.

[65] M. Kalakrishnan, P. Pastor, L. Righetti, and S. Schaal, "Learning objective functions for manipulation," in *2013 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2013, pp. 1331–1336.

[66] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning." in *AAAI*, 2008, pp. 1433–1438.

[67] A. Boularias, J. Kober, and J. R. Peters, "Relative entropy inverse reinforcement learning," in *International Conference on Artificial Intelligence and Statistics*, 2011, pp. 182–189.

[68] G. E. Hovland, P. Sikka, and B. J. McCarragher, "Skill acquisition from human demonstration using a hidden markov model," in *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on*, vol. 3. Ieee, 1996, pp. 2706–2711.

[69] D. Lee and Y. Nakamura, "Stochastic model of imitating a new observed motion based on the acquired motion primitives," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*. IEEE, 2006, pp. 4994–5000.

[70] S. Calinon, F. Guenter, and A. Billard, "On learning the statistical representation of a task and generalizing it to various contexts," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*. IEEE, 2006, pp. 2978–2983.

[71] S. Calinon and A. Billard, "Learning of gestures by imitation in a humanoid robot," Cambridge University Press, Tech. Rep., 2007.

[72] S. Calinon, "Robot programming by demonstration," in *Springer handbook of robotics*. Springer, 2008, pp. 1371–1394.

[73] D. H. Grollman and A. G. Billard, "Robot learning from failed demonstrations," *International Journal of Social Robotics*, vol. 4, no. 4, pp. 331–342, 2012.

[74] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Machine learning*, vol. 3, no. 1, pp. 9–44, 1988.

[75] G. Neumann and J. R. Peters, "Fitted q-iteration by advantage weighted regression," in *Advances in neural information processing systems*, 2009, pp. 1177–1184.

[76] J. Peters and S. Schaal, "Reinforcement learning of motor skills with policy gradients," *Neural networks*, vol. 21, no. 4, pp. 682–697, 2008.

[77] N. Kohl and P. Stone, "Policy gradient reinforcement learning for fast quadrupedal locomotion," in *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, vol. 3. IEEE, 2004, pp. 2619–2624.

[78] J. Tan, Y. Gu, C. K. Liu, and G. Turk, "Learning bicycle stunts," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 50:1–50:12, Jul. 2014. [Online]. Available: http://doi.acm.org/10.1145/2601097.2601121

[79] K. Ding, L. Liu, M. van de Panne, and K. Yin, "Learning reduced-order feedback policies for motion skills," Tech. Rep. TR-2012-06, University of British Columbia, Tech. Rep., 2012.