UNIVERSITY OF CALIFORNIA, SAN DIEGO

Computational and Psychophysical Studies of Goal-Directed Arm Movements

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Cognitive Science

by

Dan Liu

Committee in charge:

Professor Emanuel Todorov, Chair
Professor Robert Bitmead
Professor Howard Poizner
Professor Marty Sereno
Professor Virginia de Sa

2008

The dissertation of Dan Liu is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2008

To my parents.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am fortunate to have benefited from the help of many people during my time at UCSD.

First and foremost, I owe my deepest gratitude to Professor Emo Todorov for supervising the work in this thesis. His kindness, wisdom and patience are more than I could have hoped.

I am grateful to other members of my committee for their time and support. Professor Howard Poizner gave much valuable feedback on my research and has always been a delight to talk to. I am thankful to Professor Marty Sereno and Professor Bob Bitmead for their many memorable classes. Professor Sereno showed me how fascinating the brain is, and Professor Bitmead showed me how powerful optimal control is, both encouraged me to apply optimal control to understanding brain functions. Thanks are due to Professor Virginia de Sa, who has been very accommodating and showed much interest in my work. She also provided a wonderful role model of how to balance a career in science with starting a family, which keeps me from dropping out of grad school.

I am thankful to Professor John Batali and Richard Belew for thoughtful discussion and encouragement on my second-year and third-year projects.

I thank current and past students in the natural computation group and the Cognitive Science Department, for numerous fruitful discussion and support. I am especially grateful to Weiwei Li, Xiuchuan Pan, Daphne Chen, Ben Huh, Paul Hammon, Alan Robinson, Alex Simpkins, Miro Enev, Yuval Tassa, Mike Kelley, Flavia Filimon and Danke Xie.

I am indebted to the cogsci staff, particularly Beverley Walton, Elizabeth Cuevas, Marine Sinanyan, Mark Wallen and Thanh Maxwell for their excellent administrative support.

Further thanks for friendship go out to Junwen Wu, Dashan Gao, Janet Zhou, Tang Tang, Fang Fang, Jiucang Hao, Yang Chen and Cissy Liu. In particular I would like to thank Junsheng Han for his support during my thesis writing.

Last and most importantly, I thank my parents, Changfu Liu and Changyou Xu, for their love and encouragement on pursuing my dream. To them, I dedicate this dissertation.

Chapter 3, in part, was originally published in *the Journal of Neuroscience*, 27(35):9354-9368,2007. Chapter 4, in part, has been submitted for publication in *the Journal of Neuroscience*. Chapter 5, in part, has been submitted for publication in *IEEE International Symposium on*

Adaptive Dynamic Programming and Reinforcement Learning. The dissertation author was the primary researcher and author of all these papers.

VITA

2000	B.E., Automation, Tsinghua University,Beijing, China.
2003	M.S., Control Theory and Engineering, Tsinghua University, Beijing, China.
2008	Ph.D., Cognitive Science, University of California, San Diego.

PUBLICATIONS

D. Liu and E. Todorov. "Hierarchical Control as an Approximation to Optimal Control on a Realistic Arm Model", submitted to *IEEE International Symposium on Adaptive Dynamic Programming and Reinforcement Learning*, 2008.

D. Liu and E. Todorov. "Smart Learning", submitted to *the Journal of Neuroscience*, 2008.

D. Liu and E. Todorov. "Adaptation on Demand", *Society of Neuroscience (SfN)*, San Diego, Nov. 2007.

D. Liu and E. Todorov. "Learning Mechanism in Reaching Movement under Target Jump Paradigm", *Progress in Motor Control VI International Society of Motor Control*, Sao Paulo, Brazil, Aug. 2007.

D. Liu and E. Todorov. "Evidence for the Flexible Sensorimotor Strategies Predicted by Optimal Feedback Control", *The Journal of Neuroscience*, 27(35):9354-9368, 2007.

L. Lu, D. Liu and H. J. Zhang. "Automatic Mood Detection and Tracking of Music Signal", *IEEE Trans. on Speech and Audio Processing*, 14(1): 5-18, 2006.

D. Liu and E. Todorov. "Trial-to-Trial Inference in Open-Loop and Closed-Loop Adaptation", *Workshop on Bridging the Gap Between Sensation and Motor Control*, Rauischholzhausen, Germany, Jul. 2006.

D. Liu and E. Todorov. "Visuomotor Feedback Gains Adapt to the Statistics of the Environment", *ESF-EMBO Symposia on Three-Dimensional Sensory and Motor Space*, Sant Feliu de Guixols, Spain, Oct. 2005

D. Liu and E. Todorov. "Flexible Strategies for Sensorimotor Integration in Uncertain Environments", *Conference of Neural Control of Movement (NCM)*, Florida, Apr. 2005.

D. Liu and E. Todorov. "Visually Guided Corrections Elicited by Target Perturbations are Consistent with Optimal Feedback Control", *Society of Neuroscience (SfN)*, San Diego, Nov. 2004.

D. Liu, L. Lu and H.J. Zhang. "Automatic Mood Detection from Acoustic Music", *the 4th International Conference on Music Information Retrieval*, Maryland, Nov. 2003.

D. Liu and N.Y. Zhang. "A Computer Aided Design System for Developing Performance Programs for Music Fountains", *Tsinghua Science and Technology*, 8(5):612-616, 2003.

D. Liu, N.Y. Zhang and H.C. Zhu. "Form and Mood Recognition of Johann Strauss's Waltz", *Chinese Journal of Electronics*, 12(4):587-593, 2003.

ABSTRACT OF THE DISSERTATION

Computational and Psychophysical Studies of Goal-Directed Arm Movements

by

Dan Liu

Doctor of Philosophy in Cognitive Science

University of California San Diego, 2008

Professor Emanuel Todorov, Chair

Movements produced in everyday life pursue a goal. Key to the success of such movements is the motor system's ability to adjust sensorimotor strategies in a flexible way according to the goal. On the high level, flexibility entails taking into account multiple task requirements and properties of the environment and preparing a sensorimotor strategy customized for the present task and circumstances to better achieve the goal. On the low level, a strategy is flexible if it makes on-line adjustments that exploit the multiple ways in which a redundant musculoskeletal plant can achieve the same behavioral goal. Both levels of flexibility, however, are mostly ignored by traditional theories. This thesis uses both psychophysical experiments and computational modeling to explain how biological movements arise from different goals they pursue.

Our first focus is on how task goals shape motor planning. We show that the motor system customizes sensorimotor strategies for current task requirements, rather than generating a rigid motor trajectory regardless of the goal. We account for such customization of task goals in the optimal feedback control framework by using a composite cost function instead of a homogeneous cost with multiple hard constraints.

We also address how a control strategy is adapted in changing environments. We show that motor learning involves not only the statistical formation of an internal model to predict external changes, but also the flexible use of such predictions to adjust motor commands for maximum performance. Such flexible dependence on predictions is accommodated by extending the optimal control framework to deal with complicated noise and cost formulations.

Finally, rather than ignoring the musculoskeletal structures of human body, we apply

a hierarchical control framework to a more realistic arm model with 7 degrees of freedom. This framework is inspired by the facts that optimal feedback controllers for redundant systems exhibit hierarchical organization, and that sensorimotor control occurs simultaneously on many levels. The basic idea is to have the high level solve the optimal control problem with reduced dimensionality, and the low level perform an instantaneous feedback transformation of plant dynamics according to the high-level commands. This work sheds light on understanding how the brain controls human body which is a complex redundant system.

Chapter 1

Introduction

The study of motor control can make use of a large number of experimental techniques including behavioral psychophysical studies, single-cell electrophysiology in non-human primates, electromyographic (EMG) measurements of muscle activities, functional magnetic resonance imaging (fMRI) on brain activities, and transcranial magnetic stimulation (TMS) on selected brain areas. In order to combine the findings of such a large and varied collection of experimental evidence to better understand motor function on a more fundamental level, we need some notion of how the brain processes information and implements control. In other words, how the brain solves the computational problem. The main theme of this thesis is to build a quantitative model of some aspects of movement control, and use behavioral experiments to verify the modeling.

Given the high complexity of the central nervous system (CNS) and the motor apparatus itself, studying behavior from a computational perspective always requires making simplifications. For example, traditional theories consider movements generation as two separate processes. During the planning phase, a desired trajectory is generated based on some criteria; during the execution phase, such trajectory is realized using servo control mechanisms [10, 11, 12, 13, 13, 14, 15]. The appeal of this idea is that it is easy to implement. In fact, most industrial robots are controlled in this way. However, given the notion that human movement is far more graceful than that from a industrial robot, using only simple models bears the risk of missing the essence of what makes our motor systems special. Indeed, although these theories account for a large number of averaged behavioral data, they cannot explain the trail-to-trial variability, e.g., the trajectories are rarely the same when a movement is made repeatedly even

for a well-trained athlete. The challenge hence lies in building "good" models that are complex enough to capture the underlying computational principles, yet simple enough to be solvable.

We believe that optimal feedback control [4, 16, 17, 1, 6] satisfies the above criteria to be a good model in studying biological movements. First, the biological processes that continuously improve behavior closely resemble iterative optimization. This makes optimal control theory a natural framework for studying the neural control of movement. Secondly, different from traditional optimal control models that only optimize a desired trajectory and ignore on-line sensory feedback, optimal feedback control models also optimize a sensorimotor transformation (control law) to utilize the on-line sensory feedback in a flexible way to better archive the performance. Indeed, optimal feedback control models have arguably been more successful than any other classes of models in terms of explaining the details of experimental data [17]. In particular, the flexible sensorimotor transformation generated from optimal feedback control has provided a prefect tool in studying the flexible sensorimotor strategies of the motor system, which is key to the success of various motor behavior.

Previous work on optimal feedback control has emphasized the sensorimotor strategies during movement execution, in particular the structure of motor variability and the goal-directed nature of on-line corrections [18, 19, 20, 21, 4]. Such flexibility has been explained with the minimal intervention principle, which states that task irrelevant deviations from the average behavior should be left uncorrected to maximize performance [4, 17, 16]. The flexibility in motor planning/preparation, in particular the the systematic relationship between sensorimotor strategies and mixtures of task goals, on the other hand, has received surprisingly little attention. Optimal control models, which dominate the thinking on trajectory planning, have traditionally optimized a homogeneous cost and treated all other goals as hard constraints, the latter are supposed to be specified externally, outside the scope of such models. The homogeneous cost could be energy consumption [22, 23], derivative of hand acceleration [10], derivative of joint torque [11], or end-point variance [12]. The constraints include endpoint position, final velocity and acceleration (typically zero), movement time, and intermediate points along the trajectory, all of which are rarely explicit in real-world tasks, thus raising two questions: (i) how are their values being chosen; (ii) are their values "chosen" in the first place, or are they stochastic outcomes of the complex interactions among sensorimotor strategy, noise, musculoskeletal dynamics, and environment – like any other feature of individual movements? Previous analysis [4, 16] showed

that choosing desired values for movement parameters that are not explicitly specified by the task is suboptimal, no matter how the choice is made. This answers question (ii) and renders question (i) irrelevant. We show that as the relative importance of these components is varied by the experimenter, subjects modify their strategy in agreement with our theory. As in prior stochastic optimal control models [12, 4, 24], taking into account the empirically established signal-dependent nature of motor noise [25, 26, 24, 27] turns out to be important.

Another form of sensorimotor system's flexibility is its remarkable ability to adapt in a changing environment. Such adaptability has been demonstrated in adaptation experiments using visual perturbations [28, 29, 30] as well as force-fields produced by a robotic manipulandum [8, 31, 32]. The common finding is that an internal model of the experimentally imposed perturbation is acquired, and then used to generate predictive compensation [33, 34]. Despite extensive work on the statistical formulation of internal models [35, 30, 36, 37, 38], the question of how the CNS uses these internal models to produce motor commands has received surprisingly little attention. There is an implicit assumption in the literature that the output of internal models is in one-to-one correspondence with changes in motor behavior, or in other words the control system itself is not changing. This assumption has no reason to be true and indeed we show here that it can be systematically violated. In particular, we demonstrate that the same perturbation sequence can elicit very different forms of adaptation as far as motor behavior is concerned, even though the information content (and presumably the internal model being formed) is the same. These differences are found to be explained within the framework of optimal feedback control if certain novel extensions are made to the framework. The extensions have to do with robustness (as in high-order energy costs) as well as a preference for control strategies that tend to be successful in everyday life outside the lab.

The ultimate goal in the study of motor control is to understand how the brain controls the rest of the body to give rise to a large variety of skilled movements. Although optimal feedback control models are extremely powerful in explaining high-level computational principles of motor control, they are not applicable to controlling complex redundant systems such as human body due to the high dimensionality of the state and control spaces. In fact, most optimal feedback control models in explaining arm movements have to simplify the entire arm into a point mass [4, 16]. As a result, they cannot explain important questions such as how to choose one particular arm configuration, among the many alternatives, to achieve the desired

end-effector location. As an attempt to solve this control problem on redundant systems, [7] proposed a hierarchical control framework. This framework is inspired by two observations. First, from a computational viewpoint, optimal feedback controllers for redundant systems exhibit hierarchical organizations [4, 16]. Secondly, from a biological viewpoint, it is known that sensorimotor control occurs simultaneously on many levels [39, 40]. Lower-level circuits (e.g., the spinal cord) interact with the musculoskeletal system directly by both receiving rich sensory input and generating corresponding motor outputs before the rest of the brain has had time to react to that input. High-level circuits (e.g., the motor cortex), on the other hand, operate on a more abstract and goal-related movement representations [41]. The proposed hierarchical framework also has two layers. The plant is augmented with a low-level feedback controller, which receives information about the plant state, and sends to the high level more abstract representation that captures the task-relevant aspects of plant dynamics but has reduced dimensionality. The high level monitors task progress, and issues commands to archive the goal. Then the low-level controller computes energy-efficient controls to match the high-level commands. In this way, the high level solves the optimal control problem without considering all the details of the plant and thus avoid running into the curse of dimensionality, the low level performs an instantaneous feedback transformation to deal with the details. We apply this framework to a more realistic arm model with 7 degrees of freedom (DOF) and 14 muscles whose dynamics resemble those from real muscles.

1.1 Organization of Thesis

The rest of the thesis is organized as follows.

Chapter 2 reviews the literature on studies of the sensorimotor system from three aspects: neural basis, motor behavior and computational models. The goal is to raise the attention on issues that have been over-simplified by traditional computational models of biological movements, which motivate our work.

Chapter 3 focuses on the sensitivity of movements to task goals. We argue that the CNS relies on sensorimotor strategies optimized for composite cost functions based on specific task requirements, instead of optimizing a homogeneous cost and treating all other goals as hard constraints.

Chapter 4 extends optimal feedback control framework to account for motor learning in a changing environment. In particular, we examine how statistical properties of the environment is used in generating predictive movements to better achieve the behavioral goal.

As an attempt to understand how the brain controls human body, which is characterized by complex musculoskeletal properties, Chapter 5 applies a hierarchical control framework to a more realistic arm model with 7 degrees of freedom in tasks such as reaching, drawing, and orienting.

Chapter 6 summarizes the thesis.

Chapter 2

Background

2.1 The organization of movement

2.1.1 The hierarchical structure of the motor system

Skilled motor behavior is a result of an intimate interaction between the complex mechanical properties of the body and a highly distributed control system. This control system is known to be organized in a hierarchical structure: Lower levels have a variety of reflex circuits that govern the temporal details of an involving movement; high levels operates on more abstract and more goal-related movement representations. Fig. 2.1 illustrates a simplified hierarchical structure of the CNS with three levels: spinal cord, brain stem, and forebrain. The spinal cord is the lowest level, which includes motor neurons as the "final common pathway" for all motor output [42], and interneurons that integrate sensory feedback with descending commands from higher centers. This level mediates a variety of reflexes and rhythmic automatisms such as locomotion and scratching. The brain stem is the second level in the motor hierarchy, which integrates spinal reflexes into a variety of automated movements that control posture and locomotion. The highest level is provided by the cerebral cortex, which projects directly to the spinal cord and also regulates motor tracts originated in the brain stem. The most important areas in this level includes the primary motor cortex (M1), posterior parietal cortex (PPC), and the prefrontal cortex (PFC). M1 contributes the largest number of axons to the corticospinal tract and receives input from other cortical regions that are predominantly involved in motor planning. M1 neurons have the unique property of coding anything from hand directions to detailed patterns

of muscle activity [43]. PPC is crucial for internal feedback loop and thus on-line corrections. The most convincing argument is provided in [44], where the smooth on-line corrections during a reaching movement was found to be disrupted when a single transcranial magnetic stimulation (TMS) was applied over the left intraparietal sulcus(IPS) at the onset of hand movement. PPC may also play an important role in integrating sensory signals from many modalities (e.g. visual, proprioceptive, auditory and vestibular), as well as afferent copy signals from motor structures [45]. PFC is important for the organization of goal-directed, rule-based behaviors [46, 47].

In addition to this hierarchical structure, the cerebellum and basal ganglia also provide feedback circuits that regulate cortical and brain stem motor areas.

1. Cerebellum

   It is widely believed that the cerebellum regulates movement by both predicting the consequence of taking an action and providing the neural command necessary to archive some desired trajectory. The fact that cerebellum patients had difficulty adapting their motor commands when limb dynamics were altered [48] provides direct evidence for the above hypothesis. This idea is also supported by the experiments showing that although monkeys were able to learn to contract the extensor triceps muscle to prevent the elbow from overshooting after the forearm was unexpected perturbed, such predictive movement was disturbed when some deep nuclear neurons in the cerebellum were deactivated [49].

2. Basal ganglia

   Basal ganglia is usually assumed to be important in the selection and initiation of movements, the regulation of continuous movements, the control of movement sequences and of other automated actions, and the scaling of movement parameters. Both Huntington's disease (HD) and Parkinson's disease (PD) are due to pathological changes in the basal ganglia. In HD, degeneration begins in the striatal patches that inhibit dopamine (DA) neurons, resulting in an underestimation of the likelihood of success as the movements progresses. This causes the frequent change of control policies for small errors and leads to excessive movement [48]. PD, on the other hand, is due to the decreased level of dopamine in the basal ganglia. PD patients are found to be able to make on-line corrections only when the arm was not visible. This suggests that basal ganglia dependent circuits are important in integrating sensory information from different sources, especially

proprioception, and in transforming the integrated signal for motor utilization [50].

2.1.2 Areas involved in motor learning

One characteristic of the motor system is its remarkable ability of learning. Many areas in the hierarchical structure are involved in motor learning, and different areas may in change of different types of learning [2], see Fig. 2.2.

1. Cerebellum

   The cerebellum is proposed to be specialized for supervised learning based on the error signal encoded in the climbing fibers. The most direct evidence comes from Ito in vestibulo-ocular reflex (VOR) adaptation experiments, showing that long-term depression (LTD) of the Purkinje cell synapses dependent on the climbing fiber input is the neural substrate of such error-driven learning [51].

2. Basal Ganglia

   The basal ganglia is suggested to drive the variability in behavior necessary for trial-and-error learning based on reinforcement [52]. Dopamine-containing neurons have been demonstrated to encode the probability of reward (sometimes the conditional probability of reward) in their phasic firing and the uncertainty of the prediction in their maintained firing levels [53, 54]. [55] also found that the neural activity responses shifted relative to the expected reward value, and the gain adjusted to the variance of reward value.

3. Cerebral cortex

   The cerebral cortex is assumed to make unsupervised learning based on Hebbian plasticity and reciprocal connections within and between cortical areas.

2.1.3 Psychophysical principles obeyed by voluntary movements

Rather than studying the neural basis for generating movements, psychophysical studies analyzed motor behaviors during various voluntary movements and have revealed some general laws of biological movements summarized as follows:

1. Smooth movement trajectories: Trajectories of the hand between the spatial targets are relatively straight from the start to end of movement, and the velocity of the hand usually

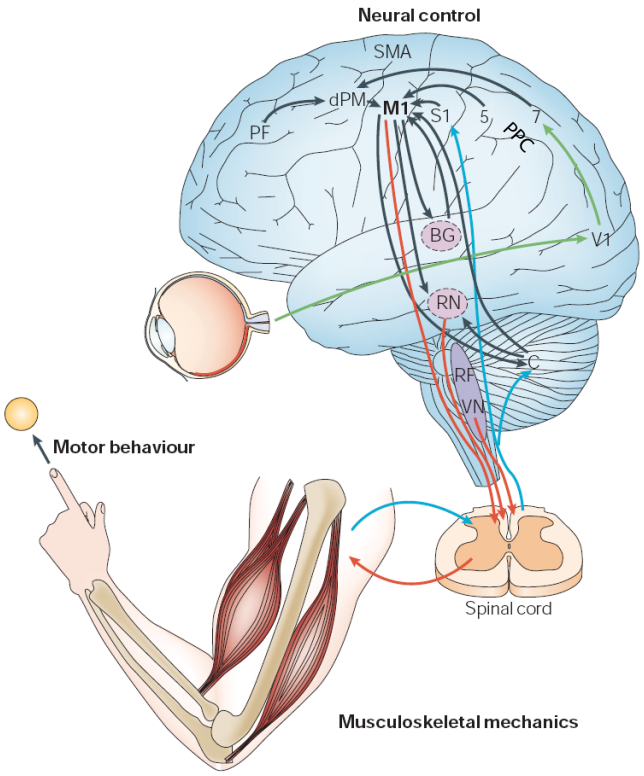
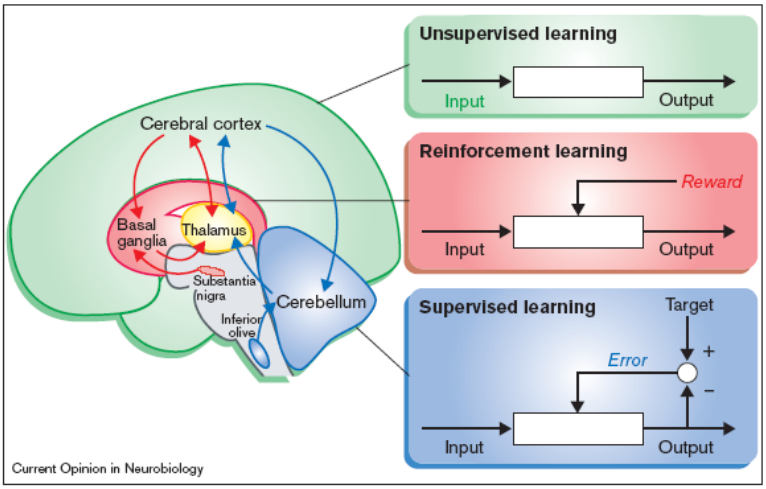
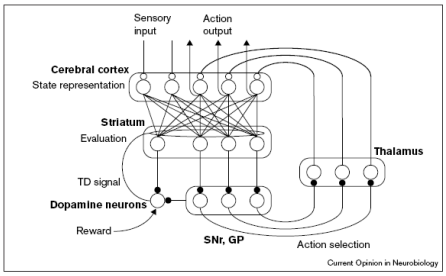


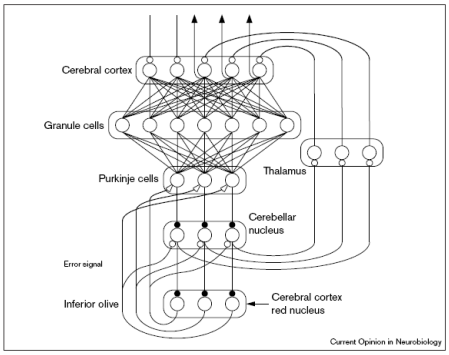
Figure 2.1 Hierarchical organization of the CNS [1]. VN, vestibular nuclei; RF, reticular formation; C, cerebellum; RN, red nucleus; BG, basal ganglia; V1, primary visual cortex; M1, primary motor cortex; PF, prefrontal cortex; dPM, dorsal premotor cortex; SMA, supplementary motor area; S1, primary somatosensory cortex; 5, parietal cortex area 5; 7, region of posterior parital cortex.



(a)



(b)



(c)

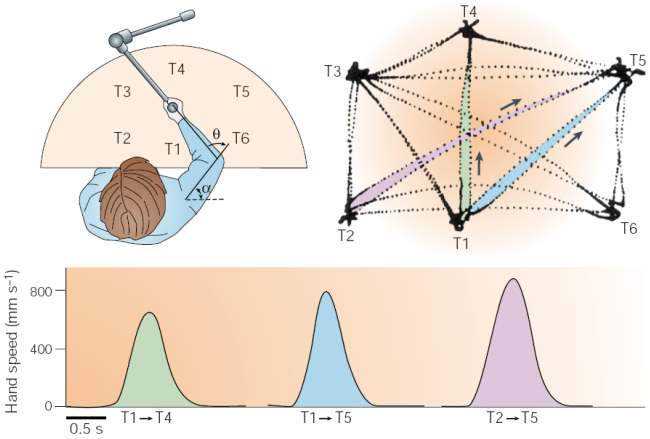
Figure 2.2 Illustration of the three main learning components [2]. **(a)**, Specialization of the cerebellum, the basal ganglia, and the cerebral cortex for different types of learning. **(b)**, Schematic diagram of the cortico-basal ganglia loop and the possible roles of its components in a reinforcement learning model. The neurons in the stratum predict the future reward for the current state and the candidate actions. The error in the prediction of future reward, the temporal difference (TD) error, is encoded in the activity of dopamine neurons and is used for learning at the cortico-striatal synapses. The filled and open circles denote inhibitory and excitatory synapse, respectively. **(c)**, Schematic diagram of the cortico-cerebellar loop. In a supervised learning model of the cerebellum, the climbing fibers from the inferior olive provide the error signal for the Purkinje cells (PC). Coincident inputs from the inferior olive and the granule cells result in LTD of the granule-to-Purkinje synapses. The filled and open circles denote inhibitory and excitatory synapses, respectively.

shows a characteristic bell-shaped profile with peak hand velocity proportional to movement distance (Fig. 2.3(a)). The motor equivalence theory [3] proposed that a purposeful movement is represented in the brain in some abstract form rather than as a series of joint motions or muscle contractions. This is consistent with the idea that the high level in the hierarchical structure represent the more abstract way, and leaves all the details to the low levels.

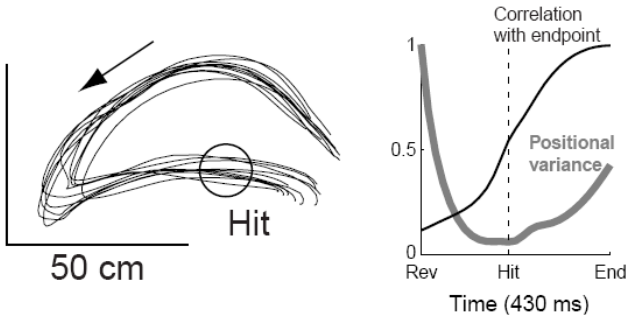
2. Relationship between path curvature and hand angular velocity : Hand movements are usually slowed down during curved movement segments and sped up during straight segments. This phenomena is captured by the two-thirds power law based on observations during drawing or scribbling. Denote path curvature with c and hand angular velocity with ω, then the two-thirds power law states $\omega \propto c^{\beta}(\beta \approx 2/3)$ or equivalently $v \propto r^{1-\beta}$ where v is the tangential velocity and r is radius of curvature $(1/\omega)$.

3. Trade-off between movement duration and terminal accuracy: Faster movements are usually less accurate for a given amplitude, captured by Fitt's law [56]. Denote movement distance with d, movement duration with t, and target width with w, then Fitt's law states $t = c_1 + c_2 \log_2(\frac{d}{w} + 1)$, where c_1, c_2 are experimentally determined constants.

4. Trial-to-trial variability: Despite the fact that movement can be remarkably precise, even a simplest reaching movement is seldom performed twice in exactly the same way. Fig. 2.3(b) illustrated the trial-to-trial variability of hand motion when subjects hit a ping-pong ball repeatedly [4]. Such variability is mainly due to the inherent noise in the sensorimotor system. For example, in an experiment to study the trial-to-trial variation of Purkinje cell (PC) activity in the floccular complex of the cerebellum during smooth-pursuit eye movement, it has been shown that some correlated variation are shared across a population of neurons and can not be canceled out by averaging. This highly covariant PC activity leads to motor variation in pursuit initiation even before the noise accompanied with motor output is sent out [57].



(a)



(b)

Figure 2.3 Features of motor behavior. **(a)**, Trajectories and velocities of the hand during point-to-point reaching movements in the horizontal plane using a mechanical linkage to monitor motion of the hand in space [3]. Trajectories of the hand between the spatial targets are relatively straight from the start to end of movement and the velocity of the hand shows a characteristic bell-shaped profile with peak hand velocity proportional to movement distance. **(b)**, Hand motions when subjects hit a ping-pong ball repeatedly [4].

2.2 Optimization principle in sensorimotor control

Section 2.1 summarized some key results regarding biological movement based on neurological and psychophysical studies. An important challenge for understanding motor functions is to connect the neural control with motor behavior by taking the complex limb mechanics into account. To do this, computational modeling provides a useful tool. Among the many computational models aiming to explain biological movement, optimal control models have been the most successful ones. The appeal of optimality principles lies in the fact that they only require a performance criterion that describes what the goal is and then fill in all movement details automatically by searching for the control strategy to best achieves the performance. Such a criterion is referred to as a 'cost function', defined as a scaler function that depending on the current set of control signals as well as the set of variables describing the current state of the musculo-skeletal system and environment. Depending on how on-line sensory feedback is used in optimization, optimal control modes can be classified into two categories: open-loop and closed-loop optimization. Open-loop optimization usually assumes deterministic dynamics and ignores the role of on-line sensory feedback. Closed-loop optimization, in contrast, assumes stochastic system and tries to construct the sensorimotor transformation (or feedback control law) that yields the best possible performance. Key models in each category will be reviewed next.

2.2.1 Open-loop optimization

Most existing optimal control models use open-loop optimization, where performing a task is assumed to have two separate stages: motor planning and motor execution. During motor planning, the best movement trajectory is formed by optimizing some cost function. During motor execution, some predefined servo mechanism is used to track the desired trajectory.

Motor planning

A fundamental property of the motor system is the redundancy. For example, to reach a target, there are many trajectories and velocities the hand may follow. Given each trajectory and velocity profile, each location along the path can be achieved with multiple combinations of joint angles. Finally, each arm configuration can be accomplished by different muscle activations, due to the overlapping actions of muscles and the ability to co-contract (Fig. 4.1). Motor planning

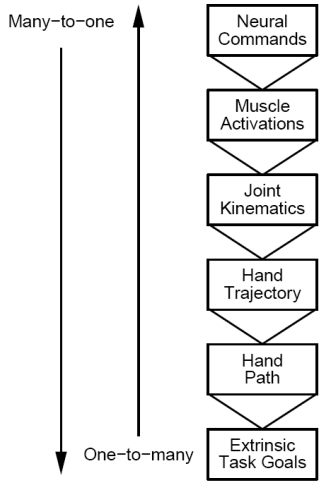


Figure 2.4 The levels in the motor hierarchy are shown with the triangles between the levels indicating the reduction in the degrees of freedom between the higher and lower levels [5]. Specifying a pattern of behavior at any level completely specifies the patterns at the level below but may result from different output from the level above. Planning can be considered as the process to choose specific patterns at each level to achieve the extrinsic task goals.

is considered as the computational process of choosing one particular pattern, among the many alternatives, to achieve the goal.

Many theories have been proposed to find the "optimal plan", differing mainly in how "optimal" is defined, or what cost function is being minimized.

1. Minimum Jerk Model [10]

   Based on the observation that point-to-point movements of the hand are always smooth in the Cartesian space, it was proposed that a major goal of motor coordination is to produce the smoothest possible movement of the hand. In this model, the cost function is defined as the square of the magnitude of jerk (derivative of hand acceleration in the Cartesian space) integrated over the entire movement.

   This model has been testified in unconstrained point-to-point movements, unconstrained curved movements, and obstacle-avoidance movements. In addition to generating smooth movements, it also accounts for the two-third power law for movement along a constrained path.

However, since this model is based solely on the kinematics of movement while neglecting the dynamics of the musculoskeletal system, it is successful only when formulated extrinsic space and when the demands of the movement lie within the capabilities of the neuromuscular system. Also, it is not clear why people want to generate smooth movement. As the author pointed out "It is not suggested that minimizing the jerk is the single objective underlying all movements. Minimization of mean-squared jerk is a mathematical model of one movement objective, the production of smooth, graceful movements".

2. Minimum Torque-change Model [11]

   Different from the minimum jerk model that considers only the geometry of movement, this model takes the dynamics of the arm into account, and generates motor commands required to achieve the movement directly. This model defines the cost function as the square of the derivative of joint torque integrated over the entire movement, based on the notion that the control minimizing the change of torque generates the smooth torque movement and therefore reduces wear and tear on the musculoskeletal system. Meanwhile, the consumption of energy is relatively low because unnecessary force is avoided.

   This model is closely related to the minimum jerk model, because acceleration is locally proportional to torque at zero speed. However, by taking into account the dynamics, the minimum torque-change model successfully predict the curved path as observed in movement under external force and movement through a sequence of via points, rather than the straight path predicted by the minimum jerk model.

   Despite its success, the minimum torque-change model, as the minimum jerk model, still does not answer the question why the CNS should optimize this quantity other than the others, how the CNS estimate such complex quantities and then integrate them over the duration of a movement, and how to select the movement duration. Most importantly, neither of the above two models can explain the motor variability.

3. Minimum Variance Model [12]

   This model defines the cost function as total positional variance in the end of the movement. This is based on the observation that neural commands have signal-dependent noise, whose standard deviation increases linearly with the absolute value of the neural control signal.

This approach has several important ramifications. First, because abrupt changes in the trajectory of the eye or arm require large driving signals which would generate more noise and therefore are suboptimal, this model explains why optimal trajectories are inherently smooth. Secondly, signal-dependent noise inherently imposes a trade-off between movement duration and terminal accuracy, consistent with Fitt's law. Thirdly, different from previous models that predicts only the average trajectories, this model also predicts the pattern of variability. Most importantly, it provides a biologically plausible theoretical underpinning for both eye and arm movements. Since such costs are directly available to the nervous system,the optimal trajectory could be learned from the experience of repeated movements

However, as an open-loop method, it ignores the on-line sensory feedback. Since variability is significantly affected by feedback especially, the predicted variance is less reliable in movements of longer duration.

Motor execution

After the motor plan is generated, sequence of the motor commands is needed to generate the desired movement trajectory and correct the error caused by the noise in both the sensorimotor system and the environment during the movement. Such process is referred to as "inverse problem". This inverse problem is difficult to solve mainly due to the fact that the human body is a highly non-linear system with huge redundancies. As the result, discovering the values of the model parameters may be time-consuming, and extra criteria is needed to select one parameter set among the many alternatives all consistent with the data. Among the many methods, the Equilibrium-Point (EP) hypothesis attracts the most attention since it takes advantage of the springlike muscle properties, and solves the inverse problem easily without considering the complicate non-linear dynamics. This idea roots in two facts: (1) a limb is at static equilibrium in the absence of external loads when all the torques generated by opposing muscles cancel out; (2) when the net stiffness due to muscle action on the limb is positive, neural actives related to the limb movement can be translated into corresponding equilibrium angle, determined by the balance of the springlike torques that keep the limb at rest [13]. According to this hypothesis, the equilibrium acts as centers of attraction, and the difference between actual and desired position generates a springlike force directed toward a virtual position, like a servo controller. Although

the early Alpha model [15] only focuses on the feed-forward path and can not deal with external force, the later Lambda model [13, 14] also considers the feedback mechanisms and therefore is able to take dynamics of the system into account. Nevertheless, the Lambda model can not explain explicitly how to control the equilibrium point to move different external loads since this gets into inverse dynamic problems again.

Criticism of the EP hypothesis usually focuses on four aspects. First, theoretical values necessary to produce very fast forearm movements are much higher than those estimated from human arm [58]. Secondly, the EP hypothesis can not explain why sensory feedback is important in precise movement [59]. Thirdly, it is not clear how to get the "equilibrium positions". Minimum Jerk model provided a way to find such desired trajectories based on the geometry of the movement. Later, this model was modified so that it could account for the trajectory modification task [60] and obstacle avoidance task [61]. Nonetheless, it is not known how such a trajectory might be computed in more complex tasks.

2.2.2 Optimal feedback control

In addition to optimizing the desired trajectory as in open-loop optimization, closed-loop optimization also constructs the best possible transformation from states of the body and environment into control signals, or the control law, to better utilize the on-line feedback in achieve the goal. As the result, it can accounts for not only the averaged behavior as open-loop optimization does, but also the trial-to-trial variability during movement. Closed-loop optimization is usually referred to as optimal feedback control [4, 16, 17, 1, 6].

General framework

The general framework of optimal feedback control is illustrated in Fig. 2.5. It is composed of both an estimator and a controller. The goal of the estimator is to integrate different sources of information to reduce the overall uncertainty and compensate for the delay, independent of the behavioral goal; whereas the goal of the controller is to determine what actions to take at each time step to best achieve the overall performance.

1. Optimal estimator

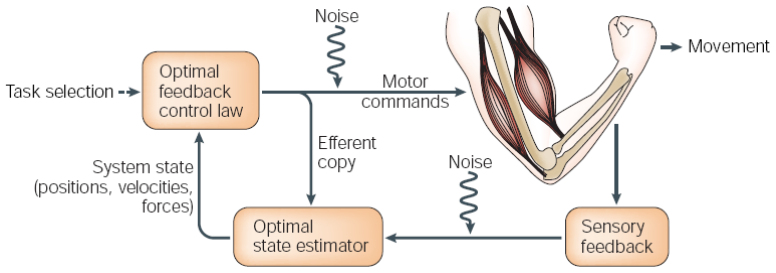


Figure 2.5 Schematic illustration of optimal feedback control [1].

Since the state of the plant (i.e. the arm in reaching movements) is observable only through delayed and noisy sensors, a recursive estimator is developed to predict internal state changes before the corresponding sensory data have arrived. This predictive capability of the estimator allows the controller to counteract disturbance before they cause errors.

Input to the estimator includes both the sensory inflow (information from visual feedback and proprioception) and the motor outflow (the motor commands sent to the arm). Evidence of using sensory feedback comes from the fact that visually guided movements are in general more accurate, and movement from deafferent patients are less accurate [62]. Evidence of using the motor outflow has been provided in a series studies showing that the self-produced tactile stimulus is perceived as less ticklish than the same stimulus generated externally [63]. Based on the fMRI result, it has been proposed that the cerebellum might be involved in predicting the specific sensory consequences of movements, providing the signal that is used to cancel the sensory response to self-generated stimulation.

The estimator is usually implemented by a Kalman filter [64]. Kalman filter an optimal estimation when the system is linear and the noise is Gaussian [64], and can be extended to accommodate non-linear system [65]. It integrates different sources of information, such as sensory data, recent control signals, earlier output, as well as the knowledge of body dynamics, in proportion to their reliability regarding the current state, and thereby reduces the overall uncertainty. Although little is known how the Kalman filter is implemented in the brain, many studies have shown that movements under different kinds of uncertainties can be explained using Kalman filter. The most well documented study is from [33], where participants were asked to estimate the location of the hand at the end of movements made

in the dark and under externally imposed forces. They found that subjects overestimated the extent of unseen arm movements, and the temporal dynamics of such estimation could be fully accounted by a Kalman filter that integrates the sensory inflow with motor outflow.

2. Optimal controller

   Based on the estimation of current states and a predefined cost function to evaluate how the current action is related to the overall performance, the controller explores all the possible controls and selects the one that minimizes the cost. Different from open-loop optimization where the control law is predefined arbitrarily, here the control law is determined by requirement of specific tasks as well as the noise and delay in the sensorimotor system.

   A key feature of the controller is its ability to choose one out of many possible solutions from the redundant system such as our motor system. This has been explained by the "minimal intervention" principle [16], which states that deviations away from the average behavior is not corrected unless those deviations interfere with task performance. This idea is explained in Fig. 4.3. The goal of the task is to use the control as small as possible to maintain $x_1 + x_2 = target$. If the variance of initial errors is a circular Gaussian, then traditional open-loop methods force $x_1 = x_2 = target/2$ and try to reduce the errors equally from all directions resulting in the gray circle. On the other hand, the optimal controller only depends on $x_1 + x_2$ but not individual values of x_1 and x_2. As the result, it pushes the states only along the task-relevant direction and leaves the error in the redundant direction uncorrected, represented by the black ellipse. According to the minimal intervention principle, trial-to-trial variability during movement is not due to sloppiness, but actually indicates a good quality of the sensorimotor system.

Solution to optimal feedback control

Optimal feedback control is in general hard to solve, except for the simple case of linear dynamics, quadratic costs and Gaussian noise sources (LQG) [66]. To model motor behavior in more realistic situations, the iterative LQG (ILQG) algorithm [67] provides a more flexible framework. It iteratively uses linearizations of the nonlinear dynamics around the accurate trajectory, and improves that trajectory by minimizing a quadratic approximation to the optimal cost-to-go function (estimated accumulated cost till the end of movement assuming the

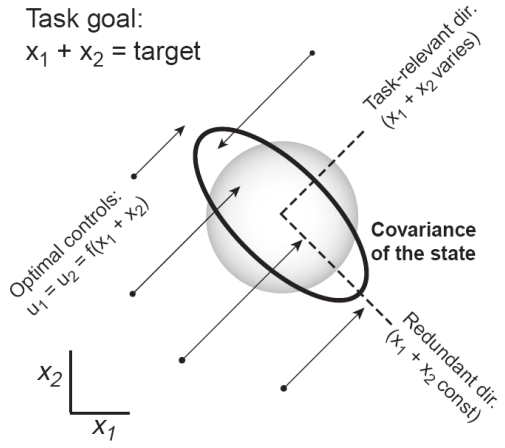


Figure 2.6 Properties of optimal feedback controllers in redundant tasks [6].

behavior is optimal). As a result, it can deal with nonlinear systems such as our arm, with more complex models of cost. In addition, it allows us to embed statistical distributions of external perturbations in the framework. Both LQG and ILQG, however, require a predefined duration, which is usually not specified in real movements. One modification is to add a term to punish long duration in the cost function [68], whereas a more general way to solve the optimal control problem is to discretized the state and control spaces and convert it into a Markov decision problem (MDP) [69] which can be solved via dynamic programming [70]. The MDP model can deal with any forms of dynamics with any forms of noise and any forms of cost functions. However, when the dimension of the states in describing the dynamical system is too big (>6), such method becomes inapplicable due to the curse of dimensionality. Despite the nice mathematical formula and wide success in explaining behavioral data of the above methods, details in the low-level of the motor system such as dynamic mechanical properties of muscle, the natural coordinates of somatosensory receptors, and the interneuronal circuitry of the spinal cord, are usually oversimplified or even ignored. In order to better capture the real control problem faced by the brain, which is much more complex than a industry robot, [7] proposed a hierarchical framework. The low levels receive rich sensory input and interact with the world directly. The high levels operate on more abstract and more goal-related movement representations with lower dimensions. In engineering models such as LQG and ILQG in modelling biological motor control, low levels

are usually oversimplified or even ignored because they include too many details such as the dynamic mechanical properties of muscle, the natural coordinates of somatosensory receptors, the interneuronal circuitry of the spinal cord, and computational noise. In order to better capture the real control problem faced by the brain, which is much more complex than a industry robot, [7] proposed a hierarchical framework for approximately optimal control of redundant manipulators such as our arm. We next review some of the mathematical foundations for LQG, ILQG and the hierarchical control.

1. Linear-Quadratic-Gaussian (LQG) [66]

   Optimal feedback control for a linear dynamical system with multiplicative noise can be modeled in discrete time as follows:

$$\text{Dynamics} \qquad \mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + \xi_t + \sum_{i=1}^{c} \varepsilon_t^i C_i \mathbf{u}_t \tag{2.1}$$

$$\text{Feedback} \qquad \mathbf{y}_t = H\mathbf{x}_t + \omega_t + \sum_{i=1}^{d} \epsilon_t^i D_i \mathbf{x}_t \tag{2.2}$$

$$\text{Cost per step} \qquad l(t, \mathbf{x}_t, \mathbf{u}_t) = \mathbf{x}_t^T Q_t \mathbf{x}_t + \mathbf{u}_t R_t \mathbf{u}_t \tag{2.3}$$

$$\text{Objective} \qquad \min_{\mathbf{u}} \sum_{t=0}^{n} l(t, \mathbf{x}_t, \mathbf{u}_t) \tag{2.4}$$

   where $\mathbf{x}_t \in \mathbb{R}^m$ is the state vector, $\mathbf{u}_t \in \mathbb{R}^p$ is the control vector, $\mathbf{y}_t \in \mathbb{R}^k$ is the feedback vector, and $t \in [0, ..., n]$ is the discrete time index. A, B, H are dynamical and observation matrices, C and D are multiplicative control and observation matrices. The mean and covariance of the initial state, as well as all the matrices are known. The control cost matrix R is symmetric positive definite, the state cost Q is symmetric positive semidefinite, and the final cost is $\mathbf{x}_n^T Q_n \mathbf{x}_n$. Noise terms ϵ_t, ω_t, ε_t, ϵ_t are modeled as independent random variables with multidimensional Gaussian distributions with mean 0 and covariances $\Omega^\xi \geq 0$, $\Omega^\omega > 0$, $\Omega^\varepsilon = I$, $\Omega^\epsilon = I$.

   Different from open-loop optimization where the goal is to find the control function based on the dynamics and previous control output, here the goal is to find $\mathbf{u}_t(\mathbf{u}_0, ...\mathbf{u}_{n-1}, \mathbf{y}_0, ...\mathbf{y}_{n-1})$ to minimize the expected total cost by taking into account the on-line feedback $\mathbf{y}_0, ...\mathbf{y}_{n-1}$

   .

Classic LQG problem only considers additive noise (i.e., $C_1, ..., C_c = 0, D_1, ..., D_d = 0$) and the solution includes a Kalman filter and a Linear-Quadratic regulator **??**.

Linear-Quadratic Regulator	Kalman Filter
$\mathbf{u}_t = -L_t\hat{\mathbf{x}}_t$	$\hat{\mathbf{x}}_{t+1} = A\hat{\mathbf{x}}_t + B\mathbf{u}_t + K_t(\mathbf{y}_t - H\hat{\mathbf{x}}_t)$
$L_t = (R + B^T S_{t+1} B)^{-1} B^T S_{t+1} A$	$K_t = A\Sigma_t H^T (H\Sigma_t H^T + \Omega^\omega)^{-1}$
$S_t = Q_t + A^T S_{t+1}(A - BL_t)$	$\Sigma_{t+1} = \Omega^\xi + (A - K_t H)\Sigma_t A^T$

The Kalman filter updates the estimated current state $\hat{\mathbf{x}}_t$ in a way to reduce the overall uncertainty. The linear-quadratic regulator, on the other hand, depends on the output from the Kalman filter and the history of control. Note the Kalman gain K does not depend on the cost or control, whereas the control law L does not depend on the noise covariance or the filter coefficients. This is because when the noise does not depend on the state or control, estimation and control can be treated separately.

However, one characteristic of human movement is that the noise scales proportional to the force. In another word, $D_1, ..., D_d$ are no longer zeros. When such multiplicative noise is present, the above independence property between estimation and control no longer holds. Assuming that the filter is non-adaptive and does not change as a function of the specific controls and observations within a simulation run, one iterative algorithm has been developed [7] which is guaranteed to converge. The algorithm starts with an initialized Kalman gains $K_0, ...K_{n-1}$ chosen arbitrarily, and then computes the control law based on this Kalman filter. Next, the Kalman filter is modified according to the updated control law and such iteration continues until it converges. Equations to compute the control law and Kalman gain are summarized in Equation 2.5 and Equation 2.6.

$$
\begin{aligned}
\mathbf{u}_t &= -L_t \hat{\mathbf{x}}_t && (2.5)\\
L_t &= (R + B^T S^{\mathbf{x}}_{t+1} B + \sum_i C_i^T (S^{\mathbf{x}}_{t+1} + S^{\mathbf{e}}_{t+1}) C_i)^{-1} B^T S^{\mathbf{x}}_{t+1} A \\
S^{\mathrm{x}}_t &= Q_t + A^T S^{\mathrm{x}}_{t+1}(A - BL_t) + \sum_i D_i^T K_t^T S^{\mathbf{e}}_{t+1} K_t D_i; && S^{\mathrm{x}}_n = Q_n \\
S^{\mathbf{e}}_t &= A^T S^{\mathrm{x}}_{t+1} B L_t + (A - K_t H)^T S^{\mathbf{e}}_{t+1}(A - K_t H); && S^{\mathbf{e}}_n = 0 \\
s_t &= \mathrm{tr}(S^{\mathbf{x}}_{t+1}\Omega^{\xi} + S^{\mathbf{e}}_{t+1}(\Omega^{\xi} + \Omega^{\eta} + K_t \Omega^{\omega} K_t^T)) + s_{t+1}; && s_n = 0
\end{aligned}
$$

The total expected cost is $\hat{\mathbf{x}}_1^T S^{\mathbf{x}}_1 \hat{\mathbf{x}}_1 + \mathrm{tr}((S^{\mathbf{x}}_1 + S^{\mathbf{e}}_1)\Sigma_1) + s_1$.

$$
\begin{aligned}
\hat{\mathbf{x}}_{t+1} &= (A - BL_t)\hat{\mathbf{x}}_t + K_t(\mathbf{y}_t - H\hat{\mathbf{x}}_t) + \eta_t && (2.6)\\
K_t &= A\Sigma^{\mathbf{e}}_t H^T (H\Sigma^{\mathbf{e}}_t H^T + \Omega^{\omega} + \sum_i D_i(\Sigma^{\mathbf{e}}_t + \Sigma^{\hat{\mathbf{x}}}_t + \Sigma^{\hat{\mathbf{x}}\mathbf{e}}_t + \Sigma^{\mathbf{e}\hat{\mathbf{x}}}_t) D_i^T)^{-1} \\
\Sigma^{\mathbf{e}}_{t+1} &= \Omega^{\xi} + \Omega^{\eta} + (A - K_t H)\Sigma^{\mathbf{e}}_t A^T + \sum_i C_i L_t \Sigma^{\hat{\mathbf{x}}}_t L_t^T C_i^T; && \Sigma^{\mathbf{e}}_0 = \Sigma_0 \\
\Sigma^{\hat{\mathbf{x}}}_{t+1} &= \Omega^{\eta} + K_t H \Sigma^{\mathbf{e}}_t A^T + (A - BL_t)\Sigma^{\hat{\mathbf{x}}}_t (A - BL_t)^T + \\
&\quad (A - BL_t)\Sigma^{\hat{\mathbf{x}}\mathbf{e}}_t H^T K_t^T + K_t H \Sigma^{\mathbf{e}\hat{\mathbf{x}}}_t (A - BL_t)^T; && \Sigma^{\hat{\mathbf{x}}}_1 = \hat{\mathbf{x}}_0 \hat{\mathbf{x}}_0^T \\
\Sigma^{\hat{\mathbf{x}}\mathbf{e}}_{t+1} &= (A - BL_t)\Sigma^{\hat{\mathbf{x}}\mathbf{e}}_t (A - K_t H)^T - \Omega^{\eta}; && \Sigma^{\hat{\mathbf{x}}\mathbf{e}}_0 = 0
\end{aligned}
$$

Here η_t is the internal noise added to the estimator to account for the multiplicative noise, which affects the estimator and thus indirectly affects the control law. The Kalman gain and the control law now depends on each other. When multiplicative noise is gone, Equation 2.5 and 2.6 are identical to the previous linear-quadratic regulator and Kalman filter. This model is different from previously introduced minimum variance model [12], since it not only optimizes the average sequence of control signals, but also the feedback gains that determine the on-line sensory-guided adjustments.

This LQG model is the first optimal control model of reaching that incorporates signal-dependent noise and combines state estimation and feedback control into an optimal sensorimotor loop, and has been applied to various reaching and eye movements [66, 71, 68].

2. Iterative Linear-Quatratic-Gaussian (ILQG) [72, 67]

When the dynamical system is nonlinear, optimal feedback control can be framed as follows:

$$\text{Dynamics} \qquad d\mathbf{x} = \mathbf{f}(\mathbf{x}, \mathbf{u})dt + F(\mathbf{x}, \mathbf{u})d\omega \tag{2.7}$$

$$\text{Cost-to-go} \qquad v^{\pi}(t, \mathbf{x}) \triangleq E[h(\mathbf{x}(\mathbf{T})) + \int_t^T l(\tau, \mathbf{x}(\tau), \pi(\tau, (x(\tau)))d\tau] \tag{2.8}$$

$$\text{Objective} \qquad \min_{\mathbf{u}} v^{\pi}(0, \mathbf{x}_0) \tag{2.9}$$

with state $\mathbf{x} \in \mathbb{R}^m$, control $\mathbf{u} \in \mathbb{R}^m$ and standard Brownian motion noise $\omega \in \mathbb{R}^p$. $l(t, \mathbf{x}, \mathbf{u}) \geq 0$ and $h(\mathbf{x}(T)) \geq 0$ are the instantaneous cost rate and final cost respectively, which do not need to be quadratic. The admissible control signal may be constrained, but the constraints are assumed to be convex. Since it is hard to find the globally optimal control law $\pi^*(t, \mathbf{x})$ independent of the initial state, the ILQG algorithm seeks locally-optimal control law based on specific $\mathbf{x}_0$.

To compute the control law, the above system is discretized with time step Δ, and the time index is represented by $k \in [0, n]$. To linearize the system dynamics and quadratize the cost function around the mean trajectory $\bar{\mathbf{x}}(t)$ and open-loop control sequence $\bar{\mathbf{u}}(t)$, the state and control are expressed in terms of deviations $\delta\mathbf{x}_k \triangleq \mathbf{x}_k - \bar{\mathbf{x}}_k$, $\delta\mathbf{u}_k \triangleq \mathbf{u}_k - \bar{\mathbf{u}}_k$. Then it has been proved that if the LQG approximation is affine in the form $\delta\mathbf{u} = \pi_k(\delta\mathbf{x}) = l_k + L_k\delta\mathbf{x}$, the corresponding cost-to-go function remains in the quadratic form

$$v_k(\delta\mathbf{x}) = s_k + \delta\mathbf{x}^T\mathbf{s}_k + \frac{1}{2}\delta\mathbf{x}^T S_k \delta\mathbf{x}$$

for all $k \in [0, n]$, and $v_k(\delta\mathbf{x})$ depends on the control $\delta\mathbf{u}_k = \pi_k(\delta\mathbf{x})$ through the term

$$a(\delta\mathbf{u}, \delta\mathbf{x}) = \delta\mathbf{u}^T(\mathbf{g} + G\delta\mathbf{x}) + \frac{1}{2}\delta\mathbf{u}^T H \delta\mathbf{u}$$

where $s, \mathbf{s}, S, g, G, H, l, L$ can be computed iteratively following some formulas. The main iterative algorithm can be summarized as follows:

1) Specify the initial open-loop control and compute the corresponding mean trajectory

$$\bar{\mathbf{x}}_{k+1} = \bar{\mathbf{x}}_k + \Delta\mathbf{f}(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)$$

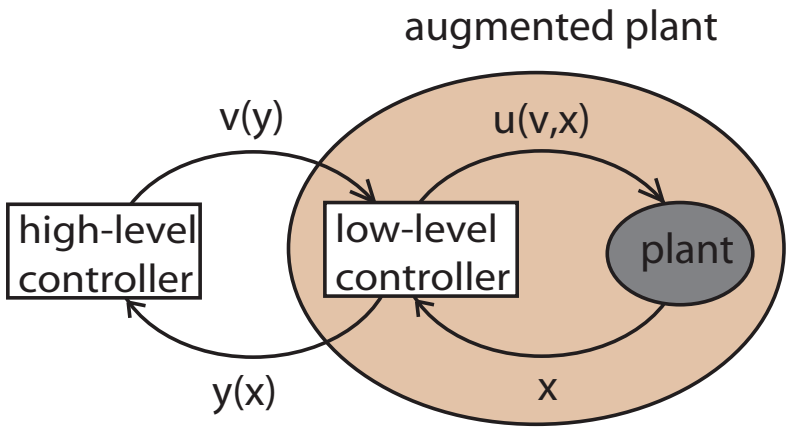


Figure 2.7 Schematic illustration of the hierarchical control framework [7].

2) Build a local LQG approximation around $\overline{\mathbf{x}}$ and $\overline{\mathbf{u}}$

3) Design an affine control law for the linearized system in the form $\delta\mathbf{u}_k = \pi_k(\delta\mathbf{x}) = l_k + L_k\delta\mathbf{x}$ and update the cost-to-go.

4) Compute the new control law $\widetilde{\mathbf{u}}_k = \overline{\mathbf{u}}_k + \delta\mathbf{u}_k$, apply the new control law forward in time to the linearized system $\delta\mathbf{x}_{k+1} = A_k\delta\mathbf{x}_k + B_k\delta\mathbf{u}_k$ initialized at $\delta\mathbf{x}_1 = 0$, and compute the new cost.

5) If the new cost and the old cost are sufficiently close, the iteration ends. Otherwise, apply Levenberg-Marquardt method in computing $\delta\mathbf{u}$ and repeat from 2. The process of Levenberg-Marquardt is as follows. When H is positive semi-definite, we have $\delta\mathbf{u} = -H^{-1}(g + G\delta\mathbf{x})$. In the case where H has negative eigenvalues (due to approximating errors), H is resembled by replacing the negative eigenvalues with a positive constant λ. If the new cost is smaller, replace $\bar{\mathbf{u}}_k$ with $\tilde{\mathbf{u}}_k$ and decrease the Levenberg-marquardt constant λ to increase the step size. Otherwise, increase λ.

3. Hierarchical framework for approximately optimal control [7]

As an attempt to control a redundant manipulator, [7] proposed a hierarchical control framework. This framework is inspired by two observations. First, from a computational viewpoint, optimal feedback controllers for redundant systems exhibit hierarchical organizations [4, 16]. Secondly, from a biological viewpoint, it is known that sensorimotor control occurs simultaneously on many levels [39, 40]. Lower-level circuits (e.g., the

spinal cord) interact with the musculoskeletal system directly by both receiving rich sensory input and generating corresponding motor outputs before the rest of the brain has had time to react to that input. High-level circuits (e.g., the motor cortex), on the other hand, operate on a more abstract and goal-related movement representations [41]. The proposed hierarchical framework also has two layers, illustrated in Fig. 5.1. The plant is augmented with a low-level feedback controller, which receives information about the plant state $\mathbf{x}$, and sends to the high-level $\mathbf{y}(\mathbf{x})$ that captures the task-relevant aspects of plant dynamics but has reduced dimensionality. The high-level monitors task progress, and issues commands $\mathbf{v}(\mathbf{y})$ to specify how $\mathbf{y}(\mathbf{x})$ should change to archive the goal. Then the low-level controller computes energy-efficient controls $\mathbf{u}(\mathbf{v}, \mathbf{x})$ to control the plant to accomplish the trajectory designed from the high-level. In this way, the high-level solves the optimal control problem without considering all the details of the plant and thus avoid running into the curse of dimensionality, the low-level performs an instantaneous feedback transformation to deal with the details.

Mathematically, the hierarchical control framework can be described as

$$\text{low-level dynamics} \quad \dot{\mathbf{x}}(t) = \mathbf{a}(\mathbf{x}(t)) + B(\mathbf{x}(t))\mathbf{u}(t) \tag{2.10}$$

$$\text{high-level dynamics} \quad \dot{\mathbf{y}}(t) = \mathbf{f}(\mathbf{y}(t)) + G(\mathbf{y}(t))\mathbf{v}(t) \tag{2.11}$$

$$\text{low-level to high-level} \quad \mathbf{y} = \mathbf{h}(\mathbf{x}) \tag{2.12}$$

where

	low-level	**high-level**
state vector	$\mathbf{x} \in \mathbb{R}^{n_x}$	$\mathbf{y} \in \mathbb{R}^{n_y}$
control vector	$\mathbf{u} \in \mathbb{R}^{n_u}$	$\mathbf{v} \in \mathbb{R}^{n_v}$
passive dynamics	$\mathbf{a}(\mathbf{x})$	$\mathbf{f}(\mathbf{y})$
control-dependent dynamics	$B(\mathbf{x})\mathbf{u}$	$G(\mathbf{y})\mathbf{v}$

Function h represents a static mapping from the low-level state $\mathbf{x} \in \mathbb{R}^{n_\mathbf{x}}$ to the high-level state $\mathbf{y} \in \mathbb{R}^{n_\mathbf{y}}$, which is selected to satisfy: (1) $\exists \tilde{q}$ s.t. $\tilde{q}(t, \mathrm{h}(\mathbf{x})) = q(t, \mathbf{x})$ so that $\mathbf{y}$ contains enough information to compute the state defendant cost, and (2) $n_\mathbf{y} < n_\mathbf{x}$ to reduce the dimensionality.

The objective of the low-level and high-level controller is as follows. The high-level defines $\mathbf{y}$ and computes the control $\mathbf{v}$ to efficiently accomplish the goal. The low-level

chooses the control $\mathbf{u}(\mathbf{v}, \mathbf{x})$ so that $\dot{\mathbf{y}}$ from the low-level will match that from the high-level, or

$$H(\mathbf{x})\mathbf{a}(\mathbf{x}) + H(\mathbf{x})B(\mathbf{x})\mathbf{u} = \mathbf{f}(\mathbf{y}) + G(\mathbf{y})\mathbf{v} \tag{2.13}$$

where $H(\mathbf{x}) = \partial\mathbf{h}(\mathbf{x})/\partial\mathbf{x}$ is the Jacobian of the function $\mathbf{h}$. When there are multiple $\mathbf{u}(\mathbf{v}, \mathbf{x})$ to satisfy this, choose the control that leads to smaller control cost $r(\mathbf{u}, \mathbf{x})$.

This framework has been applied to a 2 degreee of freedom arm model with 6 muscles during a 2D reaching task. Its validity, however, needs to be further verified using more realistic arm models on more complex tasks.

2.3 Motor learning

Everyday experience suggests that we are able to learn from a changing environment and adjust our movements accordingly. A number of studies have investigated the processes involved in motor adaptation in both predictable and randomly varying environments. The major approach is to expose subjects to specific perturbations and quantify the changes in their responses over time during a point-to-point arm reaching movements. The perturbation paradigms can be classified into three groups:

1. Mechanical perturbation

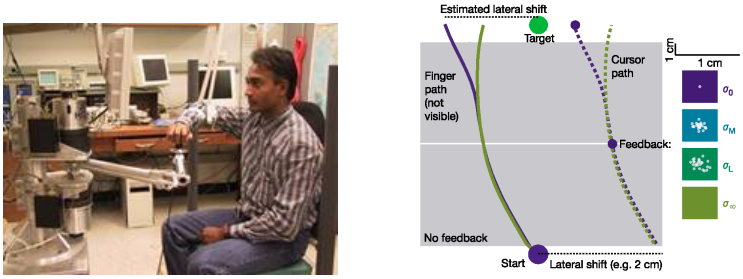
   This is to change the dynamics of the environment by perturbing the moving arm with some robot arm [8, 31, 32], see Fig. 2.8(a).

2. Visual perturbation

   This is to change the relationship between the visual inputs and motor output by displacing or deforming the visual feedback of the moving arm [28, 29, 30], see Fig. 2.8(b).

3. Target jump

   This is to change the original target (virtual or visual) position unexpectedly during a reaching movement aiming to it [73, 74, 44], see Fig. 2.8(c).



(a) (b)

(c)

Figure 2.8 Experimental paradigms for three types of perturbations. **(a)**, Experimental setup for mechanical perturbation. Subjects were asked to make arm movements while grasping the handle of a robot arm [8]. A monitor, placed directly in front of the subject and above the robot arm, displayed the location of the handle as well as targets of reaching movements. The robot arm had two torque motors at its base that allowed for production of a desired force field. Subjects were aksed to compensate for the force on the robot arm so as to bring the cursor of hand within the target square. **(b)**, Experimental paradigm for visual perturbation [9]. As the finger moved from the starting circle, the cursor was extinguished and shifted laterally from the true finger location. The hand was never visible and visual feedback was briefly displayed with different uncertainty on halfway to the target. Subjects were asked to place the cursor on the target, thereby compensating for the lateral shift. **(c)**, Experimental paradigm for target jump. Subjects were asked to move to the target, which was either stable or displaced unexpectedly during the movement. Subjects were instructed that the perturbation might occur, and asked to always move to the final target location.

2.3.1 Internal model

Many studies have reported that when the perturbation (applied to the hand, visual scene, or the target) is fixed and repeatable, people can adapt to it quite well. When the perturbations are first applied, the trajectory of the hand are distorted compared with the normal, roughly straight paths, with an increasing end error. After prolonged exposure, people learn to generate compensatory force that cancels the mechanical perturbation [8, 31, 32], adjust the moving direction of the arm to cancel the visual displacement [28, 29, 30], or move to the new target location even before it was displaced [73, 74, 44]. Then, the hand trajectories gradually become normal, straight paths, and the end errors are also decreased. When the perturbations are removed suddenly after the adaptation has occurred, the trajectories become distorted and the erroneous movement shows approximately the mirror image of the initial deviation caused by the perturbations [8, 31, 75]. This after-effect supports the idea that motor system forms an internal model of the perturbation to compensate for the delay and uncertainty of sensory feedback in a predictive fashion [33, 34]. Among the many attempts to investigate how such internal model is formed, trial-by-trial learning and Baysian learning have attracted much attention.

2.3.2 Trial-by-trial learning

Inspired by the notion that people learn from their errors, many studies investigated how errors experienced from the past affect subsequent movements. [75] first reported that the current movement is affected by only a small number of previous movements, often only one. Based on this result, [32] proposed a trial-by-trial learning model, which states that error experienced in a previous trial is used to modify the motor comments in subsequent trials to prevent similar errors from occurring. [32] also employed a simple linear dynamical system (LDS) model with a memory of one trial to quantify how such an internal model is acquired during learning. Later, [76] developed a more general class of LDS which provides a first-order approximation for any Markovian learning rule in specifying how sensory feedback on each movement affects the sensorimotor transformation in subsequent movements during motor learning.

The LDS model is represented as follows:

$$\mathbf{z}(n+1) = A\mathbf{z}(n) + B\mathbf{w}(n) + \eta(n) \tag{2.14}$$

$$\mathbf{y}(n) = C\mathbf{z}(n) + D\mathbf{w}(n) + \gamma(n) \tag{2.15}$$

where n is the trial number, $\mathbf{y}(n)$ represents the correction on trial n, $\mathbf{w}(n)$ represents the perturbation, and $\mathbf{z}(n)$ is the internal learning state free to use in whatever way needed to fit the data. $\eta(n), \gamma(n)$ are independent zero-mean random variables with covariances Q and S, representing the noise during learning and applying the learning in corrections respectively. In this model, correction $\mathbf{y}(n)$ and perturbation $\mathbf{w}(n)$ can be measured from experiments, and the goal is to find the parameter set A, B, C, D, Q, S as well as the internal learning sequence $\mathbf{z}$ to best fit the data. [76] also presented an expectation-maximization (EM) algorithm to solve this.

This model, however, has two limitations. First, although "error" was usually measured as the deviation from the baseline during the movement in most experiment [75, 32, 77], it is not clear why subjects should to care about the accuracy other than in the end of the movement during a reaching task. Also, "error" should be a relative quantity based on what the goal is, since the same amount of endpoint positional error may elicit different reactions depending on how it is related to the performance evaluation [78]. The second limitation lies in the first-order Markov assumption. Although this assumption has been demonstrated by many experiments showing that neural structures modified as a result of motor learning do not explicitly retrain memories of perturbations beyond one trial in the past [75, 32], it is against the fact that even monkeys can learn a complicate sequence of movement. Direct evidence against this assumption comes from a well designed experiment to compare the on-line corrections in different perturbation experiments. In this experiment, strong trial-by-trial learning was observed only in responds to random visual and mechanical perturbations, but not to random target jumps [77]. It has been proposed that errors in visual and mechanical perturbations are due to mis-calibration of internal models (called "execution errors") and therefore result in adaptation of internal models and subsequently changes in motor commands, whereas errors in the target jump paradigm should not change the internal model. The puzzle remains, however, whether the lack of trial-by-trial learning in the target jump paradigm was because subjects learned from the sequence of perturbations in the past and thus captured the random pattern. If this hypothesis is true, the first-order Markov assumption during motor learning no longer holds.

2.3.3 Bayesian estimation

Instead of focusing on errors in affecting subsequent movements, other studies exported how the CNS utilizes statistical properties of both the inherent noise in the sensorimotor system and the uncertainty in the world. Models based on Bayesian statistics haven been the most successful ones by taking into account sensory data, recent control signals, knowledge of body dynamics, as well as its earlier output, and weighting all these sources of information regarding the correct state in proportion to their reliability.

The most well documented study is from [36], where subjects were asked to reach to a visual target and the visual feedback of their hands were displaced with some lateral shift. On each movement, the shift was randomly drawn from a prior distribution and briefly displayed once midway through the movement. By manipulating the reliability of this visual feedback on each trial, they studied how sensory feedback and the prior knowledge about the distribution of the shift were combined. The result was explained using Maximum A Posterior (MAP) model. This model was also verified in a force field reaching experiment later [79]. In addition to explaining sensorimotor integration, Bayesian estimation has also been used to explain cue combination. For example, [80] showed that position and velocity feedback were integrated in a manner consistent with the reliability of signals in on-line control of hand movements, [35] showed that visual and haptic information was also integrated in a way depending on their relevant reliability in estimating the height of an object. [30] further suggested that during motor planning, such sensory integration depended on not only the reliability, but their relevance to the task. In a recent study [38], Bayesian estimation has been demonstrated to be able to assign errors to different timescales (i.e., fast-timescale disturbances occur when muscles fatigue, slow-timescale disturbances occur when muscles are damaged or when limb dynamics change as a result of development) and thus account for several previously puzzled learning effects, such as the initially rapid rate of adaptation followed by progressively slower rates, and the spontaneous recovery when error feedback was clamped at zero following an adaptation-extinction training episode [81]. Previously introduced Kalman filter is an optimal estimator, or Bayesian estimator, when the dynamics and sensory measurements are linear and the noise is Gaussian. A fMRI study suggests that the anterior cingularte cortex (ACC) area may conduct Bayesian integration [82].

Bayesian estimation, however, only estimates what the most likely state is currently

and predicts what the most likely state will be next, but does not address the question how the control should be generated based on the estimation. In fact, as a forward process, it can not take into account how current action will affect the states in the future and thus is unable to select the action to best archive the overall behavioral goal.

2.3.4 Learning in unstable environment

When it is impossible to construct an internal model to compensate for the externally imposed perturbations (i.e. the dynamic task is not only unpredictable but unstable), subjects were found to increase the stiffness of their arm by co-contracting muscles and thereby reduce the perturbing influence. [83] have developed a task where any deviation of the hand during the movement was exacerbated by a force acting perpendicular to the line. They showed that subjects skillfully adapted their stiffness to the instability of the environment so that stiffness increased only in the required direction and by the required amount.

Chapter 3

Evidence for the flexible sensorimotor strategies predicted by optimal feedback control

3.1 Introduction

Humans interact with a diverse and uncertain environment requiring flexible motor behavior. Here we show that the optimal feedback control theory which we [4, 16, 17, 66, 7] and others [84, 85, 86, 87, 1] have pursued affords the flexibility apparent in behavioral data, in contrast with more traditional theories [10, 11, 14, 15]. We distinguish two forms of flexibility. From the perspective of motor planning or preparation, flexibility entails taking into account multiple task requirements and properties of the environment known before movement, and preparing a sensorimotor strategy with both open-loop and closed-loop components customized for the present task and circumstances. From the perspective of motor execution, a strategy is flexible if its closed-loop component makes online adjustments that exploit the multiple ways in which a redundant musculoskeletal plant can achieve the same behavioral goal. Both forms of flexibility are obvious desiderata for a well-designed estimation-control system such as the sensorimotor system.

Prior work has emphasized the evidence for flexibility during execution, in particular the structure of motor variability (which is larger in task-irrelevant dimensions) and the goal-

directed nature of online corrections [18, 19, 20, 21, 4]. We have explained such phenomena with the minimal intervention principle – which states that task-irrelevant deviations from the average behavior should be left uncorrected in order to maximize performance [4, 17, 16]. This argument is further advanced here by showing that: (i) target displacement can cause correction before the hand has cleared an intermediate obstacle – ruling out the imaginary via-points postulated by alternative models; (ii) end-point variability matches the shape of an elongated target and feedback corrections in the redundant dimension are suppressed – as the minimal intervention principle predicts. Apart from these findings, however, our emphasis here is on flexibility in motor planning/preparation.

With the exception of speed-accuracy trade-offs [88], the systematic relationship between sensorimotor strategies and mixtures of task goals (as well as properties of the environment) have received surprisingly little attention. Optimal control models – which dominate the thinking on trajectory planning – have traditionally optimized a homogeneous cost and treated all other goals as hard constraints; the latter are supposed to be specified externally, outside the scope of such models. The homogeneous cost could be energy consumption [22, 23], derivative of hand acceleration [10], derivative of joint torque [11], end-point variance [12]. The constraints include endpoint position, final velocity and acceleration (typically zero), movement time, intermediate points along the trajectory. However these hypothetical constraints are rarely explicit in real-world tasks, raising two questions: (i) how are their values being chosen; (ii) are their values "chosen" in the first place, or are they stochastic outcomes of the complex interactions among sensorimotor strategy, noise, musculoskeletal dynamics, and environment – like any other feature of individual movements? Our previous analysis [4, 16] showed that choosing desired values for movement parameters that are not explicitly specified by the task is suboptimal, no matter how the choice is made. This answers question (ii) and renders question (i) irrelevant. Instead of satisfying self-imposed constraints, we propose that the CNS relies on sensorimotor strategies optimized for composite cost functions. In the present experiments the relevant cost components encourage energetic efficiency, endpoint positional accuracy (measured as bias and variance), endpoint stability (defined as bringing the movement to a complete stop), and movement speed (avoiding the time-out errors incurred when duration exceeds a threshold). We show that as the relative importance of these components is varied by the experimenter, subjects modify their strategy in agreement with our theory. As in prior stochastic optimal control models

[12, 4, 24], taking into account the empirically established signal-dependent nature of motor noise [25, 26, 24, 27] turns out to be important.

Even though our focus is on effects taking place before movement, the effects in question correspond to changes in a control strategy with both open-loop and closed-loop components – which in turn is best studied using perturbations. Perturbing the target of a reaching movement in an unpredictable direction has been a productive paradigm for investigating the mechanisms of online visuo-motor corrections [73, 74, 89]. Most prior studies have introduced perturbations around the time of movement onset, and found that the hand path is smoothly corrected to reach the displaced target – in agreement with multiple models of motor control [90, 60, 91, 92]. However, perturbations introduced late in the movement may be more informative because they are not fully corrected – in contradiction with alternative models, and, somewhat paradoxically, in agreement with optimal feedback control. Such phenomena have been observed with both visual target perturbations [93] and mechanical limb perturbations [94]. In case of limb perturbations, the correction reflects both neural feedback and musculoskeletal impedance – which are seamlessly integrated [95] and difficult to disentangle. Therefore we focus on target perturbations. We first design a 2D reaching experiment to rule out the trivial explanation that the incomplete correction is simply due to lack of time. We then replicate the phenomenon in our model, and find that it reflects a previously unknown trade-off between endpoint accuracy and stability. This yields a novel prediction: if stability requirements are decreased then accuracy should increase. The prediction is confirmed in 3D obstacle avoidance and interception experiments. The latter experiments give rise to rich motor behavior, allowing us to make a number of additional observations consistent with our theory. These include shaping variability patterns to buffer noise in redundant dimensions; adjusting movement duration to take advantage of temporal error margins; exploiting target impedance and surface friction to achieve end-point stability; reallocating corrective action among redundant actuators to balance signal-dependent noise and inertial constraints; and correcting for task-relevant perturbations before having reached the task-irrelevant subgoals hypothesized by alternative models.

3.2 Materials and Methods

3.2.1 Experimental Setup

Experiment 1

Seven subjects made planar reaching movements on a table positioned at chest level. A 21" flat-screen CRT was mounted above the table facing down, and was viewed in a see-through horizontal mirror. In this way computer-generated images could be physically aligned with the hand workspace. Movement kinematics were recorded with an Optotrak 3020 infrared sensor at 100Hz. A small pointer, which had an Optotrak marker and an LED attached near its tip, was held in the dominant right hand. The task was to move the LED to a starting position, wait for a target to appear, and move to the target when ready. Movement onset was detected online using a 1cm threshold on the distance between the pointer and the starting position. Analysis of speed profiles (Fig. 4.1(b)) revealed that the actual movement started about 100msec before the distance threshold was reached. Therefore we define the origin of the time axis to be 100msec before the online detection of movement onset, and report all times relative to this corrected origin.

The end of the movement was defined as the first point in time when the hand speed had remained below 0.5cm/s for 40msec. The LED was turned off at movement onset, turned on at the end of the movement, and remained on in the repositioning phase. The room was dark. The target was always visible. Thus movements were made without visual feedback of the hand, although subjects could see their endpoint error as soon as the movement ended. Movement duration was required to be between 600msec and 800msec. If the duration on any trial fell outside these boundaries, the computer displayed a "slow down" or "speed up" message respectively. Movement amplitude was 30cm. The main movement was in the lateral direction from right to left (although in Fig. 4.1(a) and Fig. 4.2(a) we plot the movements from left to right, for consistency with the space-time plots).

After a brief familiarization session every subject performed 240 trials. Within each trial the target could either remain stationary or jump 5cm forward or backward, orthogonal to the main movement direction. Subjects were instructed that jumps may occur, and asked to always move to the final target position and stop there within the allowed time interval. The jumps occurred at 100msec, 200msec or 300msec. There were 180 perturbed trials (30 for

every possible latency-direction combination) and 60 non-perturbed trials, presented in random permutation order.

Experiment 2

Eight subjects made 3D arm movements around a horizontal obstacle while aiming for the center of a vertical target (Fig. 4.4(a)). Movement dimensions are illustrated in Fig. 4.4(b). The target was a 20cm x 5cm wooden board with a bull's-eye pattern, and was mounted on a 3DOF robot (Delta Haptic Device, Force Dimension Inc). Subjects held in their right hand a 7cm wooden pointer with an electromagnetic Polhemus Liberty sensor attached to it. The sensor measured 3D position and orientation (at 240Hz) making it possible to compute the position of the tip of the pointer. The latter is referred to as the "hand". Another sensor was attached to the target (which was also tracked via the robot's encoders at 1000Hz). Before each trial the robot moved the target away and waited for the subject to initiate the trial – by inserting the tip of the pointer in a small receptacle mounted below the obstacle and remaining stationary for 100msec. Then the robot "presented" the target by moving towards the subject, at which time the subject was free to move when ready. In case of an anticipation error the computer played a sound and aborted the trial. Hand movement onset was detected with a 1cm position threshold on the distance from the starting position. Subsequent analysis of speed profiles revealed that the movement started about 50msec before it was detected. Therefore we define the origin of the time axis to be 50msec earlier.

Movement end was detected when the hand speed remained below 10cm/s for 40msec, or when the target was displaced due to impact with the hand by more than 0.4 cm. The maximum allowed movement duration was 900msec. Time-out errors were signaled by moving the target away and playing a loud sound. During the hand movement the target was either stationary, or was rapidly displaced by the robot 9cm left or right. The robot trajectory is illustrated in Fig. 4.4(c)(f); it was generated by a model-based controller with both open-loop and closed-loop components. The target jump could be initiated at 50msec (early) or 350msec (late).

After brief familiarization subjects performed 20 trials without perturbations, followed by two experimental conditions/sessions with perturbations, 120 trials each. Each experimental session included 40 early jumps, 40 late jumps, and 40 baseline trials presented in random permutation order. Left and right jumps were equally probable. Subjects were instructed that

jumps may occur, and asked to always move to the final target position within the allowed time interval. The two experimental sessions differed in the stopping requirements of the task. In the stop condition subjects were asked to slow down their hand movement before contact and touch the target gently, so that the impact would not displace the target by more than 0.4cm. If it did, the computer played an explosive sound and the target was moved away rapidly. This indicated to the subject that the target has been hit too hard. In the hit condition hitting the target hard was no longer an error. The robot was placed in high-gain servo mode (with carefully chosen non-linearities to avoid instability) and was able to absorb the impact with the hand. The maximum force output of the robot is 25N. Subjects were not explicitly asked to hit the target harder, but they quickly discovered the benefits of using such a strategy. Half of the subjects performed the hit condition first and then the stop condition; the order was reversed for the other half of the subjects.

Experiment 3

Ten subjects performed a task similar to Experiment 2, with the following modifications. A pressure sensor (FSR, Interlink Electronics) was installed inside the starting position receptacle and used to detect movement onset earlier and more reliably. An ATI Mini-40 six-axis force-torque sensor (2000Hz sampling) was mounted behind the target. It allowed more reliable detection of contact (which was now defined as the movement end) as well as direct measurement of contact force. Subjects still initiated the movement when ready. As soon as hand movement was detected, the robot began to move downward at a constant speed (6.67cm/s). This motion continued until the trial ended due to hand-target contact, or until the target hit the horizontal edge of a wooden board mounted underneath. This happened 900msec after movement onset, and was defined as a time-out error. The downward motion was repeatable and easily predictable – providing subjects with an explicit representation of allowed movement duration. When the target jumped (9cm left or right) the rapid lateral motion was superimposed on the slow downward motion. Instead of the bull's-eye pattern the target now had a pattern of vertical stripes (5cm x 1cm each), with gray levels increasing with lateral distance from the center stripe (which was white). Subjects were asked to make contact with the target as close to the center stripe as possible. In the stop condition the threshold for hitting too hard was now defined in terms of force rather than displacement (0.8N in the first 8msec after contact). The bookshelf obstacle

from Fig. 4.4(a) was now replaced with a horizontal bar, and moved backwards to induce a more curved hand movement (dimensions are illustrated in Fig. 4.5(c)). Early jumps were triggered at movement onset; late jumps were triggered at 400msec after movement onset; allowed movement duration was 900msec. Every subject still participated in two conditions – hit and stop – in counterbalanced order. In each session we scheduled 25 baseline trials, 25 early jumps, and 25 late jumps in random permutation order. Any failed trials (time-out errors or hitting-hard errors) were now rescheduled at a random time later in the same session, yielding a more balanced database of analyzable trials. Instead of 20 no-perturbation trials before the experiment, we now had 10 trials before and 10 trials after the experiment. These were used to compute variability in the absence of perturbations ("baseline" in Fig. 3.8).

In experiments 2 and 3 the wrist was immobilized with an orthopedic brace so as to avoid corrective movements using the wrist. There were two reasons for this restriction. First, pilot experiments revealed different involvement of the wrist in early versus late corrections (Fig. 4.6(c)) making the comparison between conditions difficult. Second, our models assume point-mass dynamics and do not capture wrist movements.

3.2.2 Statistical analysis

In experiments 2 and 3 both the time-out errors and the hitting-hard errors were signalled immediately, in a way that disrupted the behavior, and therefore error trials could not be included in the analysis. Only no-error trials were analyzed. They constituted 79% percent of all trials in experiment 2 and 58% in experiment 3. The higher overall error rate in experiment 3 is because we repeated failed trials, and so subjects performed more trials in the more difficult conditions (late/stop in particular). In experiment 1 errors were signalled by the computer only after the movement had stopped. Thus the error signals could not disrupt the behavior, allowing us to include trials whose duration was slightly over the time limit (up to 100msec). Presumably these movements were generated by the same underlying mechanism and the longer duration was simply due to trial-to-trial variability. 93% of all trials in experiment 1 were analyzed.

All statistical tests were based on n-factor ANOVA ("anovan" in the Matlab Statistics Toolbox). We avoided averaging to the extent possible. In the comparisons of undershoot and duration (Fig. 4.1(d)(g) and Fig. 4.4(e)(g)(h)) and wrist contribution (Fig. 4.6(c)), individual trials were treated as repeated measures, and the factors were the experimental conditions

(perturbation time, stop vs. hit when applicable) as well as the subject identity. Thus we had two factors in experiment 1 and three factors in experiments 2, 3 and the pilot experiment in Fig. 4.6(c). The subject identity was modeled as a factor with random effects since subjects are drawn randomly from the population. In the comparisons of standard deviations (Fig. 4.3(e) and Fig. 4.6(a)), time-out error rates (Fig. 4.1(h)), and lateral velocities (Fig. 4.6(b)), all trials that a subject performed in a given condition were combined to obtain a single number. All comparisons of means were based on Tukey's criterion for post-hoc hypothesis testing. Differences are reported as significant when $p < 0.05$. The error bars shown in the figures correspond to ± 1 pooled standard error of the mean, as computed by the multiple comparison function used to perform Tukey's test ("multcompare" in the Matlab Statistics Toolbox).

3.2.3 Computational models

Optimal feedback control model (LQG)

We model the hand as an $m = 1$kg point mass moving in a horizontal plane, with viscosity $b = 10$Ns/m approximating intrinsic muscle damping. The point mass is driven by two orthogonal force actuators which can both pull and push (approximating two pairs of agonist-antagonist muscles). The actuators act as muscle-like first-order low-pass filters of the control signals, with time constant $\tau = 0.05$s. These settings of m, b, τ were chosen to be compatible with biomechanics and were not adjusted to fit the data.

Let $\mathbf{p}(t), \mathbf{v}(t), \mathbf{a}(t), \mathbf{u}(t)$ be the 2D hand position, velocity, actuator state, and control signal respectively. The corresponding units are m, m/s, N, N. The time index is $t \in [0, t_f]$. The final time t_f is specified (taken from the experimental data in Fig. 4.1(g)). The plant dynamics in continuous time are modeled as

$$\begin{aligned} d\mathbf{p}(t) &= \mathbf{v}(t)\, dt \\ m\, d\mathbf{v}(t) &= (\mathbf{a}(t) - b\mathbf{v}(t))\, dt \\ \tau\, d\mathbf{a}(t) &= (\mathbf{u}(t) - \mathbf{a}(t))\, dt + M(\mathbf{u}(t))\, d\mathbf{w}(t) \end{aligned}$$

$\mathbf{w}(t)$ is standard Brownian motion. $M(\mathbf{u}(t))$ represents control-multiplicative or signal-dependent motor noise, and is given by

$c_1 = 0.15$ corresponds to 2D noise in the same direction as the control vector $\mathbf{u}(t)$, while $c_2 = 0.05$ corresponds to 2D noise in the direction orthogonal $\mathbf{u}(t)$. The parallel noise

component is larger because muscles pulling in the direction of net muscle force are more active and therefore more affected by signal-dependent noise. The parameters c_1, c_2 as well as the sensory noise magnitude σ described below are adjusted so that the baseline variability predicted by the model (Fig. 4.2(e)) is similar to the experimental data (Fig. 4.1(e)). Note that the shape of these curves cannot be fully captured by three scalar parameters, and so the good fit mostly reflects the quality of the model.

Denoting the target position $\mathbf{p}^*(t)$, we can assemble all variables into an 8D state vector

$$\mathbf{x}(t) = [\mathbf{p}(t); \mathbf{v}(t); \mathbf{a}(t); \mathbf{p}^*(t)]$$

and write its dynamics in general first-order form

$$d\mathbf{x}(t) = (A\mathbf{x}(t) + B\mathbf{u}(t))\,dt + C(\mathbf{u}(t))\,d\mathbf{w}(t)$$

with A, B, C obtained from the above equations.

In order to define an optimal control problem we also need a cost function. As in our previous work [4], we use a mixed cost function defined as

$$\|\mathbf{p}^*(t_f) - \mathbf{p}(t_f)\|^2 + w_{\text{stop}}\left(\|\mathbf{v}(t_f)\|^2 + \|s_{\text{a}}\mathbf{a}(t_f)\|^2\right) + w_{\text{energy}}\int_0^{t_f}\|\mathbf{u}(t)\|^2\,dt$$

The three cost terms encourage endpoint positional accuracy, stopping at the target, and energetic efficiency respectively. The activations $\mathbf{a}(t)$ are scaled by $s_{\text{a}} = 0.1$ because their numerical values turn out to be an order of magnitude larger than positions and velocities. The weight $w_{\text{energy}} = 0.00005$ of the control term is a free parameter (it is not clear how to estimate this parameter independently). The weight w_{stop} determines the relative importance of coming to a complete stop (i.e. achieving zero velocity and acceleration) at the end of the movement. We use $w_{\text{stop}} = 1$ for the stop condition and $w_{\text{stop}} = 0.01$ for the hit condition. These values are chosen to capture the qualitative differences between the stop and hit conditions. Note that we do not model the 3D experiments explicitly.

The state of the plant $\mathbf{x}(t)$ is not directly observable, but has to be inferred from noisy observations whose time-integral $\mathbf{y}(t)$ satisfies

$$d\mathbf{y}(t) = \mathbf{x}(t)\,dt + G d\mathbf{z}(t)$$

The sensory noise covariance is diagonal: $G = \sigma \mathrm{diag}\left(1, 1, 1, 1, 1/s_{\mathrm{a}}, 1/s_{\mathrm{a}}, 0, 0\right)$ with $\sigma = 0.015$ adjusted to reproduce the observed movement variability. $\mathbf{z}\left(t\right)$ is standard Brownian motion. In comparison with our earlier model [4], the present model is simpler in that here we use first-order rather than second-order muscle filters and do not explicitly represent sensory delays. We do not model explicitly the visuo-motor delays or uncertainty in detecting the target observations. To obtain correct reaction times, we simply model each target perturbation as occurring 120msec later than the corresponding experimental perturbation.

With these definitions, we discretize the time axis (with 1msec time step) and obtain a discrete-time linear-quadratic-Gaussian optimal control problem. The reason for formulating the model in continuous time and then discretizing the time axis, as opposed to working in discrete time all along, is that in a discrete-time formulation the model parameters are affected by the time step. If one were to change the time step it would not be obvious how the model parameters should scale. Continuous-time formulations have the advantage of being independent of discretization time steps. For details on how to discretize a continuous-time system see [96].

The presence of signal-dependent noise complicates matters, however we have derived an efficient algorithm for solving such problems elsewhere [66]. That algorithm is applied here to yield a modified Kalman filter for computing the optimal state estimate $\widehat{\mathbf{x}}\left(t\right)$, and an optimal feedback controller of the form

$$\mathbf{u}\left(t\right) = -L\left(t\right)\widehat{\mathbf{x}}\left(t\right)$$

Once the filter and controller are available, the state is initialized with the experimentally-defined starting position and $\mathbf{v}\left(0\right) = \mathbf{a}\left(0\right) = 0$, and the system is simulated until the final time t_f.

The time-varying matrix of feedback gains $L\left(t\right)$ is 2x8 and is in principle described by 16 numbers. However in the present problem it turns out to have a lot of structure which can be captured by only 3 independent parameters. In particular, the control law can be written as

$$\mathbf{u}\left(t\right) = k_p\left(t\right)\left(\mathbf{p}^*\left(t\right) - \widehat{\mathbf{p}}\left(t\right)\right) - k_v\left(t\right)\widehat{\mathbf{v}}\left(t\right) - k_a\left(t\right)\widehat{\mathbf{a}}\left(t\right)$$

where k_p, k_v, k_a are time-varying scalar gains illustrated in Fig. 4.3(a)(c).

When a perturbation is introduced in the model, the final time t_f is adjusted according to Fig. 4.1(g) and the optimal estimator and controller for the remainder of the movement are recomputed (given the new final time and target position). This is necessary because the optimal feedback gains are originally scheduled up to the duration of the unperturbed movement.

Table 3.1 Parameters of linear-quadratic-Gaussian model

parameters	**criteria**
$m = 1\text{kg},\ b = 10\text{Ns/m},\ \tau = 0.05\text{s}$	compatibility with biomechanics
$s_{\text{a}} = 0.1$	order-of-magnitude normalization
$c_1 = 0.15,\ c_2 = 0.05,\ \sigma = 0.015$ $w_{\text{energy}} = 0.00005$	fit to variability of unperturbed trials overall fit to data
$w_{\text{stop}} = 1 \rightarrow 0.01$	qualitative predictions regarding stop vs. hit

Consequently the predicted feedback corrections are not generated by exactly the same feedback gains as shown in Fig. 4.3(a)(c). However the change due to the recomputation is small, and so our intuitive analysis of feedback gains is valid. Similar re-computation is involved in the minimum-jerk feedback controller. A more general optimal feedback control model capable of predicting the changes in movement duration is described later.

The model parameters, and the criteria for choosing their values, can be summarized as follows:

Modified minimum-jerk model

The original minimum-jerk model [10] postulates that the hand moves from a starting position $\mathbf{p}_0$ to a target position $\mathbf{p}^*$ along a trajectory which minimizes the time-integral of the squared jerk (third derivative of position):

$$\min_{\mathbf{p}(\cdot)} \int_0^{t_f} \|\dddot{\mathbf{p}}(t)\|^2 \, dt$$

To make this optimization problem well-posed one has to specify the velocity and acceleration at the endpoints. Let $\mathbf{v}_0$ and $\mathbf{a}_0$ be the initial velocity and acceleration (possibly non-zero) and suppose the final velocity and acceleration are 0. Then the constraints are

$$\mathbf{p}(0) = \mathbf{p}_0,\ \dot{\mathbf{p}}(0) = \mathbf{v}_0,\ \ddot{\mathbf{p}}(0) = \mathbf{a}_0,\ \mathbf{p}(t_f) = \mathbf{p}^*,\ \dot{\mathbf{p}}(t_f) = 0, \ddot{\mathbf{p}}(t_f) = 0$$

One can find the solution to this minimization problem using the calculus of variations [10]. The expression for the optimal trajectory $\mathbf{p}(t)$ is a somewhat complicated function of $t, t_f, \mathbf{p}_0, \mathbf{v}_0, \mathbf{a}_0$, and $\mathbf{p}^*$. Differentiating that function with respect to t three times yields

$$\dddot{\mathbf{p}}(t) = \frac{60}{(t_f - t)^3}(\mathbf{p}^* - \mathbf{p}_0) - \frac{36}{(t_f - t)^2}\mathbf{v}_0 - \frac{9}{(t_f - t)}\mathbf{a}_0$$

where $t_f - t$ is the remaining movement time. The rationale for working with third derivatives is that we are given all derivatives up to second order at the initial time, and if have a way of computing the third derivative at each time, we could simply integrate and obtain the entire trajectory. This suggests a feedback-control formulation [91] in which $\dddot{\mathbf{p}}$ is only computed at the current time and is treated as an instantaneous control signal $\mathbf{u}$:

$$\mathbf{u}(t) = \dddot{\mathbf{p}}(t) = \frac{60}{(t_f - t)^3}(\mathbf{p}^*(t) - \mathbf{p}(t)) - \frac{36}{(t_f - t)^2}\dot{\mathbf{p}}(t) - \frac{9}{(t_f - t)}\ddot{\mathbf{p}}(t)$$

The three scalar coefficients in the above expression are time-varying feedback gains for a third-order system with state vector $\mathbf{x}(t) = [\mathbf{p}(t); \dot{\mathbf{p}}(t); \ddot{\mathbf{p}}(t)]$. Note that the feedback-control formulation allows us to make the target position time-varying.

In the absence of perturbations the trajectory predicted by this modified minimum-jerk model [91] is identical to the prediction of the original minimum-jerk model [91]. The advantage of the modified formulation is that it can generate feedback corrections and thus serve as a model of perturbation experiments. Note that the modified minimum-jerk model reflects a very different philosophy compared to the original model, because a trajectory plan for the rest of the movement is no longer needed. In that sense it is closer to our optimal feedback control model [4]. The empirical success of jerk-minimization has often been interpreted as evidence for trajectory planning. Such interpretations are unjustified given that the same predictions can be made without the assumption of trajectory planning.

Optimal feedback control model (MDP)

Here we describe a different optimal feedback control model where movement duration is no longer predefined and task constraints are enforced more explicitly. It is constructed using the more general but less efficient methodology of Markov Decision Processes (MDP): the continuous state and action spaces are discretized [69] and the resulting discrete optimization problem is solved via dynamic programing [70]. Discretization methods suffer from the

curse of dimensionality and only apply to low-dimensional problems. This necessitates a simplification of the dynamics model: the arm is now modeled as a fully-observable second-order plant with state vector containing hand position $\mathbf{p}$ and velocity $\mathbf{v}$ and control vector $\mathbf{u}$ corresponding to hand acceleration. All quantities are expressed in units of centimeters and seconds. The initial state is $\mathbf{p}(0) = \mathbf{v}(0) = [0; 0]$. The default target position is $\mathbf{p}^* = [20; 0]$ but can be perturbed to either $[20; 5]$ or $[20; -5]$. Instead of perturbing the target, we perturb the hand in the opposite direction (without changing hand velocity) and then correct the hand and target positions in the subsequent analysis. In this way the target can be treated as constant and omitted from the state vector.

Each trial ends when the horizontal hand position exceeds 20cm (i.e. the hand reaches the target plane) or when the duration exceeds a maximum allowed duration of 0.6sec, whichever comes first. Let t_f denote the duration of a given trial. The total cost to be minimized is defined as

$$\text{total\_cost} = \text{final\_cost} + w_{\text{energy}} \int_0^{t_f} \|\mathbf{u}(t)\|^2 \, dt$$

The final cost, computed at the end of the movement, is defined as

$$\text{final\_cost} = \begin{cases} \|\mathbf{p}^* - \mathbf{p}(t_f)\|^2 + w_{\text{time}} t_f, & \text{if } t_f < 0.6 \text{ and } \|\mathbf{v}(t_f)\| < v_{\max} \\ 100, & \text{otherwise} \end{cases}$$

The endpoint velocity threshold $v_{\max}$ is 5cm/sec in the stop condition and 20cm/sec in the hit condition. These values were chosen to match the observed endpoint velocities. The main movement amplitude (20cm), perturbation amplitude (5cm), maximum movement duration (0.6sec), constraint violation cost (100) and other simulation parameters described below were chosen in advance and were not adjusted. The only parameters which were adjusted to fit the data were $w_{\text{energy}} = 0.00003$ and $w_{\text{time}} = 20$. This was done by solving the problem multiple times for different points in $w_{\text{energy}} - w_{\text{time}}$ space. The qualitative pattern of results shown in Fig. 3.7 and 3.8 depended weakly on w_{time} but was sensitive to w_{energy}. The latter parameter, denoted r in other papers, has proven to be important in almost every optimal control model we have ever constructed.

The sizes of the discretization grids were 101 x 61 points in position (20 x 12cm, step size 0.2cm), 25 x 25 points in velocity (80 x 80cm/sec, step size 3.33cm/sec), 11 x 11 points in acceleration (1666.67 x 1666.67cm/sec^2, step size 166.67cm/sec^2), and 31 points in time (0.6sec,

step size 0.02sec). These numbers were carefully balanced so that the grid density was sufficient to allow accurate approximation, the grid range was sufficient to cover the optimal state-control trajectories, and yet the number of grid points was not intractably large. The noise was uniform and additive, and perturbed the hand velocity by up to ± 2 grid points in each time step.

The feedback control law is in the general form

$$\mathbf{u} = \pi\left(\mathbf{p}, \mathbf{v}, t\right)$$

where $\mathbf{u}, \mathbf{p}, \mathbf{v}, t$ are constrained to the corresponding grids. Unlike the LQG framework where the function π is linear and can be represented with a small number of feedback gains, here we do not know in advance the form of π. Instead we represent it as a lookup table which specifies the value of $\mathbf{u}$ for every possible combination of $\mathbf{p}, \mathbf{v}, t$. This table consists of about 220 million numbers computed by the dynamic programming algorithm in about half on hour of CPU time. To speed up the computation and be able to explore the effects of various parameters, we also implemented the algorithm on an nVidia GeForce 8800 GTX videocard with 128 parallel processors. This reduced the running time of the algorithm by about a factor of 30. Since the videocard only supports single-precision floating point arithmetic, we used it for model exploration and run the final model on an Intel CPU with double precision. Once the control laws for the stop and hit conditions were obtained (the only difference being the value of $v_{\max}$), we applied them to the stochastic plant and simulated 3000 movement trajectories per condition: 1000 without perturbation, 1000 with perturbation at 0.1sec, and 1000 with perturbation at 0.3sec.

Even though perturbations were applied in the testing phase to characterize the response of the optimal feedback control laws, the control laws themselves were optimized for an environment without perturbations. To study possible adaptation effects, we computed a second pair of feedback control laws optimized for a perturbed environment. In the latter environment the target could jump either up or down (at 0.2sec) or remain stationary. The three types of trials had equal probability. Target perturbations were taken into account in the optimization process by incorporating an appropriate position noise term (with tri-modal distribution) in the dynamics model.

3.3 Results

3.3.1 Undershoot in reaching to perturbed targets

We define "undershoot" as endpoint error in the direction in which the target was displaced. In reaching movements, undershoot (or incomplete correction) for late target perturbations has already been demonstrated [93]. However a key question remains unanswered: is the effect simply due to lack of time, or is there a more subtle reason? Experiment 1 was designed to rule out the first possibility. Subjects made lateral reaching movements on a horizontal table, without vision of the hand, while the target was displaced in an orthogonal direction (forward or backward relative to the subject) at different times during movement. Reach duration was experimentally controlled so as to ensure that even the latest perturbation could have been fully corrected if that was the only objective of the underlying control strategy. More precisely, the remaining time after the onset of the latest correction was substantially larger than the time necessary to make the same movement in isolation.

Fig. 4.1(a) shows the average hand paths for different perturbation times as well as for baseline (unperturbed) movements. Note the undershoot for 300msec perturbations. In the rest of the analysis the backward-perturbed trials are mirrored around the horizontal axis and pooled with the corresponding forward-perturbed trials. Fig. 4.1(b) shows the tangential speed profiles. The early correction is incorporated so smoothly that its effect on the speed profile is hardly visible. The late correction, on the other hand, causes a clear deviation from the bell-shaped baseline profile. One could interpret this as a discrete sub-movement superimposed on the main movement; however we will see below that the same effect can arise from a continuous optimal controller. The corrective movement, defined as movement in the forward direction, is shown in Fig. 4.1(c). The undershoot for the 300msec perturbation is significantly larger than the undershoot for the 200msec and 100msec perturbations (Fig. 4.1(d)). Note that subjects are moving without visual feedback of the hand, and therefore some misalignment between vision and proprioception [97] should be expected. This may be the cause for the slight overshoot in the 100msec conditions (indeed no overshoot was observed in the remaining experiments which were performed with visual feedback). Such misalignment should not depend on the time of the target perturbation, and so the comparison between conditions is meaningful. There is also some systematic endpoint error in the lateral direction although it shows a weaker and opposite trend

(Fig. 4.1(d)); we return to it later.

Fig. 4.1(f) shows the acceleration profile of each corrective movement, aligned on the time when forward acceleration first exceeds 5% of peak forward acceleration. The correction for the late perturbation lasts around 400msec – which is more than sufficient to make a movement with the amplitude needed for complete correction. Indeed, an accurate 5cm movement to a 1cm-diameter target should take a little under 300msec according to Fitts' law [88]. Thus the lack of complete correction is not simply due to lack of time. Yet it has something to do with time: the overall movement duration was significantly increased in late-perturbation trials, all the way up to the 800msec time limit (Fig. 4.1(g)), and consequently the percent time-out errors was increased (Fig. 4.1(h)). Although we asked subjects to treat the time limit as a hard constraint, they treated it on equal footing with the instruction to reach as close as possible to the center of the target and found a balance between these two task requirements. It is reasonable to assume that if we had convinced subjects to avoid time-out errors at all cost the undershoot would have been even larger.

3.3.2 Optimal feedback control versus alternative models

The undershoot phenomenon is inconsistent with all prior models of motor control we are aware of. One such model [60] is an extension of the minimum-jerk model of trajectory planning [10] to the domain of feedback corrections. It postulates that the hand tracks a planned minimum-jerk trajectory, and if the target is displaced, another minimum-jerk trajectory connecting the original and displaced target positions is added vectorialy to the original plan. Naturally that model predicts full correction in all cases (Fig. 4.2(f)). A related model [91], discussed in more detail below, is a feedback-control version of the minimum-jerk model. In our task it makes the same predictions as the additive model [60] in Fig. 4.2(f). Another incompatible class of models consists of equilibrium-point control [14, 15] as well as other schemes [90, 91, 92] in which the hand is drawn to the target by some virtual spring. In such models stopping can only occur when the hand reaches the target – which is in general contradiction with systematic endpoint errors that cannot be attributed to sensory-motor misalignment. In addition to the undershoot studied here, phenomena that are problematic for these models include the undershoot of primary saccades [98], the overshoot of rapid wrist movements [99, 100], and the lack of equifinality (or failure to reach the target) in certain adaptation paradigms [101, 102].

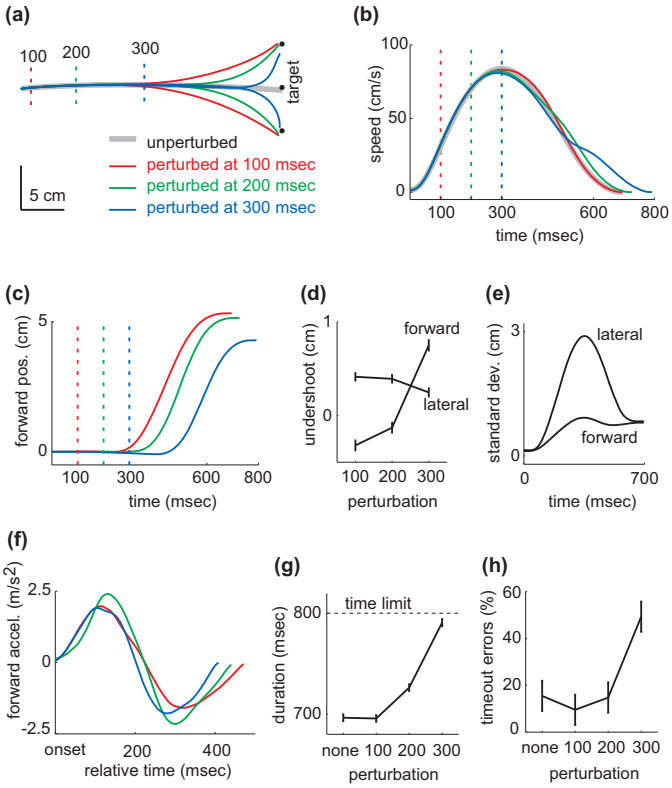


Figure 3.1 Results in Experiment 1. **(a)**, Average hand paths in experiment 1. Vertical marks show where the hand was at each perturbation time. Trajectory averaging was done as follows. The trajectory data from each individual trial were smoothed with a cubic spline ("csaps" function in the Matlab Spline Toolbox, smoothing parameter 0.001), and resampled at 100 points equally spaced in time. Analytical derivatives of the cubic spline were also computed at these 100 points – yielding velocities and accelerations. The resampled data were averaged separately in each condition. **(b)**, Tangential speed profiles for the hand paths shown in (a). **(c)**, Corrective (forward) movement. The backward-perturbed trials have been mirrored around the horizontal axis, and pooled with the corresponding forward-perturbed trials. The color code is the same as given in the legend in (a). **(d)**, Undershoot, defined as endpoint error in the direction indicated in the plot. Standard errors are computed as described in Methods. **(e)**, Positional variance of the hand trajectories in unperturbed trials. Variances at each point in time are computed separately for each subject (from the resampled data), then averaged over subjects, and the square root is plotted. **(f)**, Acceleration in the forward direction. For each perturbation time, the corresponding curve is aligned on the time when forward acceleration reached 5% of peak forward acceleration. **(g)**, Movement duration. **(h)**, Percent time-out errors, as signalled during the experiment. Note that for data analysis purposes we increased the threshold on movement duration by 100msec.

Our optimal feedback control model [4], which we have previously used to explain a number of unrelated phenomena, turns out to be compatible with the undershoot. One might have thought (as we did) that an optimal feedback controller should make larger online corrections and get closer to the displaced target than any other controller. This is not so, for interesting reasons explained in the next section. Here we only present the simulation results. The model uses a linear approximation to arm dynamics. This is justifiable because the detailed nonlinearities of the arm are unlikely to influence the response to visual perturbations significantly (they are much more relevant when it comes to resisting mechanical perturbations). The model incorporates signal-dependent motor noise [25, 26, 12, 24] as well as sensory noise, and optimizes a mixed cost encouraging endpoint positional accuracy, endpoint stability (stopping), and energetic efficiency.

The model predictions (Fig. 4.2(a)-(e)) are plotted in the same format as the data (Fig. 4.2(a)-(e)). Note the close correspondence, and in particular the undershoot for the late perturbation. When a target perturbation occurs in the model, we increase the remaining movement time as in the experimental data (Fig. 4.1(g)) and recompute the optimal controller (see Methods). If we use the unmodified optimal controller which always ends the movement at the same time, the predicted undershoot is greatly increased (dashed lines in Fig. 4.2(c)). Thus, increasing movement duration is essential for avoiding a much larger undershoot in late perturbations – which may be why subjects were so reluctant to finish the movement on time as instructed.

In Fig. 4.2(d) we also see endpoint error in the lateral direction, but its magnitude decreases with increasing perturbation time – as in the experimental data. This is because the target is not perturbed in the lateral direction, and yet the overall movement time is increased, so the control costs which cause this endpoint error (energy consumption, and reduced accuracy due to signal-dependent noise) are effectively smaller. Note also the secondary speed bump in Fig. 4.2(b) which could be mistaken for a discrete sub-movement.

In addition to reproducing average behavior, the model faithfully captures the positional variability pattern of the hand trajectories in unperturbed trials (Fig. 4.1(e) and Fig. 4.2(e)). The larger variability in the lateral (main movement) direction reflects signal-dependent motor noise – which is larger in actuators that are more active. The reduction seen towards the end of the movement is an example of structured movement variability consistent with the minimal intervention principle [4]. Another manifestation of signal-dependent noise is the increased

variability of the undershoot for late perturbations (Fig. 4.3(e)). This phenomenon is observed in the model and all experiments, and is only present in the corrective movement direction. While any corrective movement incurs signal-dependent noise in that direction, the feedback controller has less time to suppress it when the noise is introduced late.

3.3.3 Analysis of feedback gains and new predictions

The above simulation results show that the optimal thing to do is make an incomplete correction. Why is this seemingly paradoxical strategy optimal? Such questions are often meaningless: the solution to a complex optimization problem is what it is, and the relation between the problem and its solution does not have to be intuitive. Nevertheless, analysis of the model yields intuitive answers here. Key to our analysis is the fact that the optimal feedback gains are time-varying.

As explained in Methods, the optimal feedback controller can be written as

$$\mathbf{u}(t) = k_p(t)\left(\mathbf{p}^* - \widehat{\mathbf{p}}(t)\right) - k_v(t)\,\widehat{\mathbf{v}}(t) - k_a(t)\,\widehat{\mathbf{a}}(t)$$

where k_p, k_v, k_a are the optimal feedback gains, $\widehat{\mathbf{p}}, \widehat{\mathbf{v}}, \widehat{\mathbf{a}}$ are the optimal estimates of hand position, velocity, and muscle activation state (obtained by a modified Kalman filter), $\mathbf{p}^*$ is the target position, and $\mathbf{u}$ is the optimal control signal. The optimal feedback gains are illustrated in Fig. 4.3(a). We see that the positional gain k_p peaks early and then decreases in the last phase of the movement. In that phase the velocity gain k_v as well as the activation gain k_a – which can be thought of as force feedback – are large. While these gain fluctuations are hard to understand quantitatively, qualitatively they have a simple interpretation: near the end of the movement the optimal controller enters a regime where it is less sensitive to positional errors and instead aims to stop the movement in a stable manner. In retrospect this is not surprising. If we think of a mass-spring-damper system, a large spring constant (or positional gain) will make the system underdamped and cause oscillations – which is in conflict with the requirement to stop. The optimal controller effectively makes the system overdamped, achieving stability while compromising its ability to fully respond to last-minute positional errors. Thus our analysis has uncovered a trade-off between endpoint stability and positional accuracy.

Another intuitive explanation for the undershoot observed in the model is the control cost associated with large and rapid last-minute corrections. Large control signals are penalized

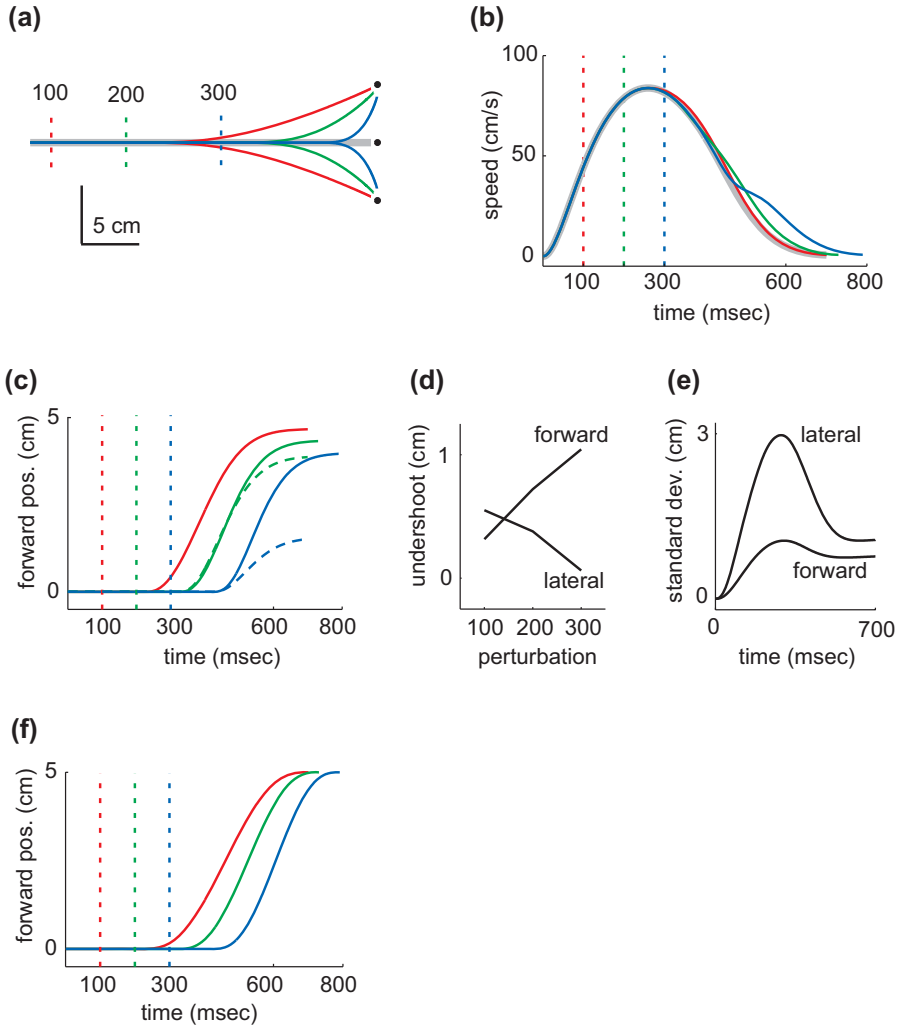


Figure 3.2 An optimal feedback control model for Experiment 1. **(a)-(e)**, Same as the corresponding subplots of Fig. 4.1, but for data generated by our optimal feedback control model. The dashed lines in (c) show predictions of a different optimal control model, where movement duration is not adjusted when a perturbation arises. There is no dashed line for the 100msec perturbation (red) because in that condition subjects did not increase the movement duration. **(f)**, Corrective movements predicted by the modified minimum-jerk model.

in two ways. One is a direct energy cost; the other is an indirect accuracy cost resulting from the signal-dependent nature of motor noise. Increased noise is particularly undesirable near the end of the movement when the feedback loop no longer has time to correct for it (Fig. 4.3(e)). In agreement with the latter interpretation, the undershoot predicted by the model increases when either the energy cost or the signal-dependent noise magnitude are increased (results not shown).

Analysis of feedback gains is also illuminating with regard to the modified minimum-jerk model and its failure to predict the undershoot (Fig. 4.2(f)). Consider the following feedback-control formulation [91] of the minimum-jerk model. At each point in time a new minimum-jerk trajectory is formed, starting at the current hand position, velocity and acceleration, and ending at the target with zero velocity and acceleration. The initial portion of this trajectory is used to control the movement, and then the procedure is repeated – making it possible to correct for online disturbances. More precisely, the hand is treated as a third-order system where the position $\mathbf{p}$, velocity $\mathbf{v}$ and acceleration $\mathbf{a}$ are state variables, and the control signal $\mathbf{u}$ is defined as the derivative of acceleration (or jerk). It can be shown (see Methods) that the minimum-jerk feedback controller has the same general form as the optimal feedback controller, but with different feedback gains:

$$k_p(t) = \frac{60}{(t_f - t)^3}, \quad k_v(t) = \frac{36}{(t_f - t)^2}, \quad k_a(t) = \frac{9}{(t_f - t)}$$

We now see something which is in retrospect obvious: the only way the minimum-jerk feedback controller can always make a full correction, regardless of how late the perturbation arises, is to use infinite feedback gains at the end of the movement. As the time t approaches the final time t_f all three feedback gains go to infinity, with k_p increasing faster than k_v and k_a. Note that we could apply this minimum-jerk controller to a partially-observable system, where state estimates are obtained by a Kalman filter, and obtain a control scheme which overall is very similar to optimal feedback control. The only important difference is in the sequence of time-varying feedback gains being used. The optimal feedback gains (Fig. 4.3(a)) not only predict behavior which better corresponds to the experimental data, but also guarantee minimal expected cost, and are finite rather than infinite (which is more biologically plausible).

We now return to the stability-accuracy trade-off, and use the model to obtain novel predictions reflecting this trade-off more directly. In the above analysis the reason for the reduced sensitivity to positional errors was the need to stop at the target. What would happen if the

importance of stopping decreases relative to the importance of reaching the target? In the model stopping is enforced with a cost term quadratic in the final velocity and activation. If we scale down this cost term, then the optimal feedback gains change as shown in Fig. 4.3(c). Note that the positional gain k_p now peaks much later. Consequently the predicted undershoot is almost eliminated (Fig. 4.3(b)). The optimal controller for the modified cost function takes advantage of the relaxed stopping requirement, and no longer brings the velocity to zero at the specified final time – particularly for late perturbations (Fig. 4.3(d)).

3.3.4 Experimental confirmation of model predictions

The above predictions were tested in experiment 2 which compared two conditions: asking subjects to stop at the target versus allowing them to hit the target. Subjects made 3D movements around a horizontal obstacle and aimed for a physical target attached to a 3D robot (see Methods and Fig. 4.4(a)). The obstacle was introduced in order to increase movement duration (so that we no longer had to impose a lower limit) and also to test the different predictions of optimal feedback control and alternative models with regard to obstacle avoidance (see next section). On randomly chosen trials the robot rapidly displaced the target, left or right, either 50msec or 350msec after movement onset. Exceeding the maximum allowed duration (900msec) resulted in a time-out error. Average movement trajectories are shown in Fig. 4.4(b).

Each subject was now tested in two conditions. In the stop condition subjects were required to slow down their movement and touch the target gently. The robot used a low-gain servo controller so that the target could be easily displaced by the hand; a displacement larger than 0.4cm resulted in a "hitting-hard" error. In the hit condition subjects were allowed to hit the target, although they were not instructed to do so. The robot used a high-gain servo controller and could absorb the impact with the hand; displacing the target no longer resulted in an error. The difference in target impedance made the distinction between the two conditions more ecologically valid.

The experimental results confirmed our model predictions. The undershoot for late perturbations was still present (Fig. 4.4(c)(d)) but it was significantly smaller in the hit condition compared to the stop condition. For early perturbations the undershoot was smaller (compared to late perturbations) and the difference between the hit and stop conditions was not significant. As before, late perturbations caused an increase in movement duration (Fig. 4.4(e)), and a sub-

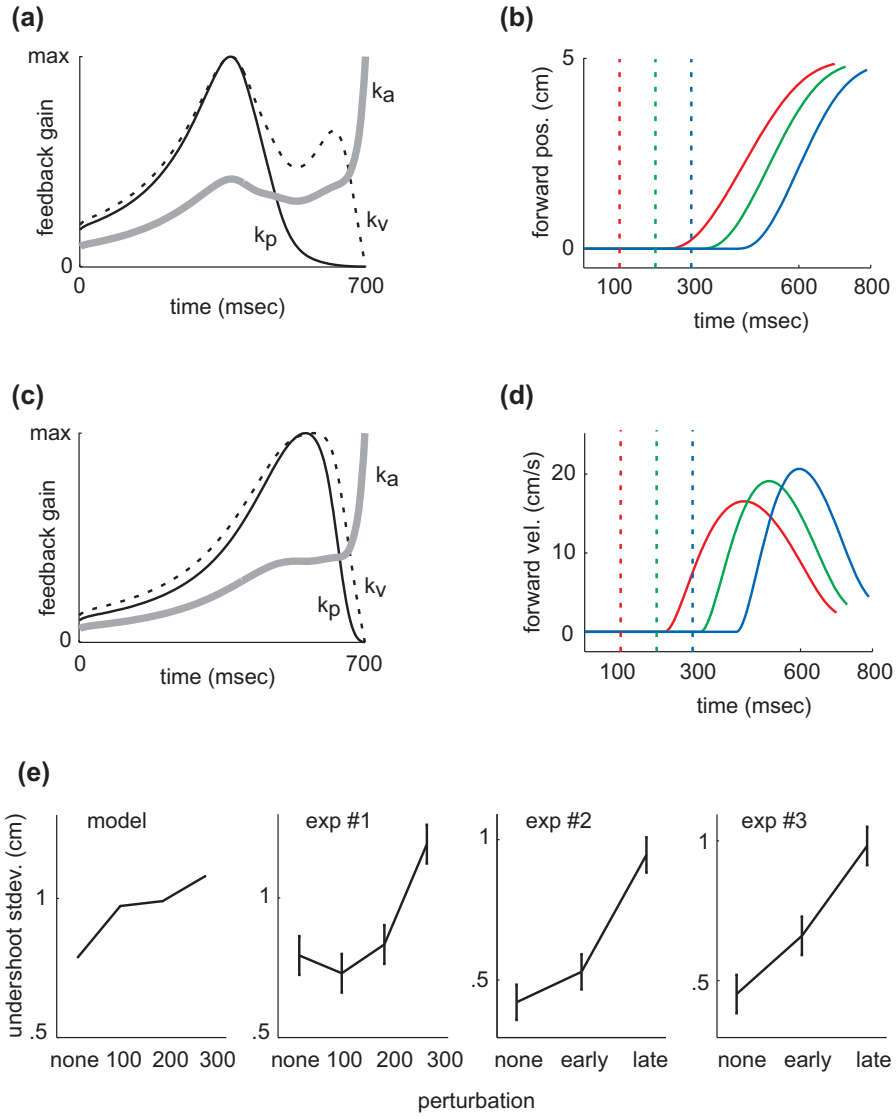


Figure 3.3 Optimal feedback gains and undershoot from Experiment 1-3. **(a,c)**, Optimal feedback gains, each scaled by its maximum value. The stop condition is shown in (a); the hit condition is shown in (c). **(b)**, Corrective movements predicted by the optimal feedback controller in the hit condition. **(d)**, Velocity of the corrective movements predicted in the hit condition. Note that velocity is not reduced to zero at the end of the movement, especially for the 300msec perturbation. **(e)**, Standard deviation of the undershoot in the model and all three experiments. The standard deviation was computed separately for each subject and perturbation time, and then averaged over subjects (by the anova procedure – see Methods). In unperturbed trials ("none"), we compute variability along the perturbation axis for the corresponding experiment, even though these trials are unperturbed.

stantial percentage of time-out errors in late/stop trials (40%). Movement duration in the stop condition was larger compared to the hit condition, and yet the correction was smaller (i.e. the undershoot was larger). Thus, as in experiment 1, subjects could have made a larger correction in the stop condition if that was their only objective.

Although at this point we had a convincing story, we remained puzzled by subjects' reluctance to treat the time limit as a hard constraint – even though in experiment 2 we made time-out errors much more salient. We reasoned that this may be because the time limit does not correspond to any physical property of the environment, and instead is signalled by the computer on the basis of an (invisible) timer. Could the outcome change if we provided an explicit and ecologically valid time cue? More importantly, could such a time cue reduce the uncertainty in the task and somehow enable subjects to eliminate the undershoot? These issues were addressed in experiment 3 where we used an interception task rather than a pointing task (see Methods). The main change was that as soon as hand movement was detected, the robot began to move the target downward at a low constant speed. Lateral target jumps were superimposed on this downward motion. Subjects were instructed to make contact with the target before it hit the horizontal edge of a board mounted underneath. The downward motion was repeatable and easily predictable, providing an explicit representation of allowed movement duration.

The results from experiment 3 (Fig. 4.4(f)(g)(h)) were similar to experiment 2, and in agreement with our model predictions. The undershoot in the stop condition was again larger than in the hit condition; the difference was now significant even for early perturbations (perhaps because we modified the method for detecting hitting-hard errors, making the threshold effectively smaller). Movement duration was again increased in late-perturbation trials, and was larger in the stop condition compared to the hit condition. The time-out error rate in late/stop trials was reduced to 31%, indicating that the explicit time cue had an effect, but this rate was still higher than what would be expected if subjects treated the time limit as a constraint. The time-out error rate in late/hit trials was much lower (12% in experiment 2 and 7% in experiment 3).

3.3.5 Absence of imaginary targets in obstacle avoidance

Optimal feedback control differs from most alternative models in that it does not invent arbitrary subgoals – such as desired trajectories or imaginary targets – but instead uses all

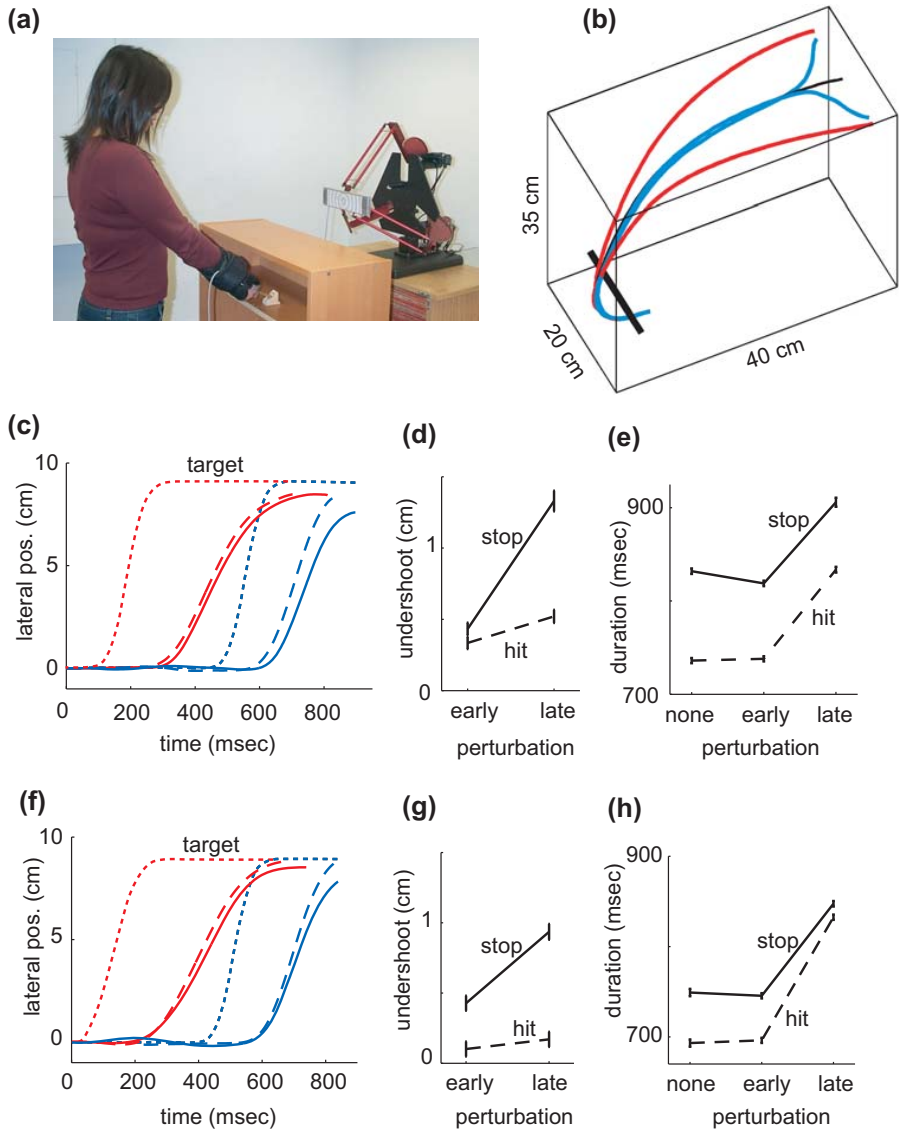


Figure 3.4 Experimental setup and results in Experiment 2-3. **(a)**, Setup for Experiment 2. Subjects make a movement from the starting position receptacle to a target attached to the robot, while clearing a horizontal obstacle (bookshelf). The robot may displace the target by 9cm left or right during the movement. **(b)**, Average hand paths in the stop condition of experiment 2. Trajectory averaging was done in a way similar to experiment 1, except that we now used a zero-phase-lag 4th-order Butterworth filter. The color code is the same as before: black – baseline; red – early perturbation; blue – late perturbation. **(c)**, Corrective movements in experiment 2. Dashed lines – hit condition; solid lines – stop condition. **(d)**, Undershoot in Experiment 2. **(e)**, Movement duration in experiment 2. **(f)**, Corrective movements in Experiment 3. **(g)**, Undershoot in Experiment 3. **(h)**, Movement duration in experiment 3.

available resources to pursue the high-level movement goal. In obstacle avoidance tasks it predicts that the hand should clear the obstacle without aiming for a specific imaginary target to the side of the obstacle. In contrast, deterministic trajectory planning models [10, 11] as well as other models [103] need such imaginary targets in order to avoid obstacles (and make curved movements in general). Note that stochastic optimal control models can avoid this limitation, by taking into account the probability of collision due to random deviations from the average trajectory [104].

Here we present two lines of evidence that subjects do not use imaginary targets in obstacle avoidance. First we analyze the variability pattern of the hand paths in unperturbed trials in experiments 2 and 3. If subjects were aiming for an intermediate target, their hand paths should be less variable in the vicinity of that target – as we have shown previously with real targets [4]. The variability pattern is plotted spatially in Fig. 4.5(a)(c) , and as a scalar quantity (variability per dimension) in Fig. 4.5(b)(d). For both experiments, and for both the hit and stop conditions, we see that the pattern is bell-shaped. In particular there is no evidence for a reduction of variability in the middle of the movement where the imaginary target should be.

Second we analyze the onset of the lateral correction relative to the time when the hand clears the obstacle and starts moving towards the robot. If subjects were aiming for an imaginary target in order to clear the obstacle, that target should be close to the reversal point and should not move when the robot displaces the final target. Therefore the corrective movement should not start before the reversal point. We focus on experiment 3, which was specifically designed to address this question by shifting the obstacle further away from the starting position (thus delaying the reversal) and detecting the onset of hand movement with a pressure sensor (allowing an earlier perturbation). Fig. 4.5(e) shows that the reversal occurs at around 300msec in both the hit and stop conditions, while the corrective lateral acceleration in early perturbations starts at 100-150msec. Given the filtering of the musculoskeletal system, the neural command driving the correction must have been generated even earlier. Thus, subjects begin to correct for the target jump before having reached the hypothetical via-point – casting serious doubt on the existence of the latter.

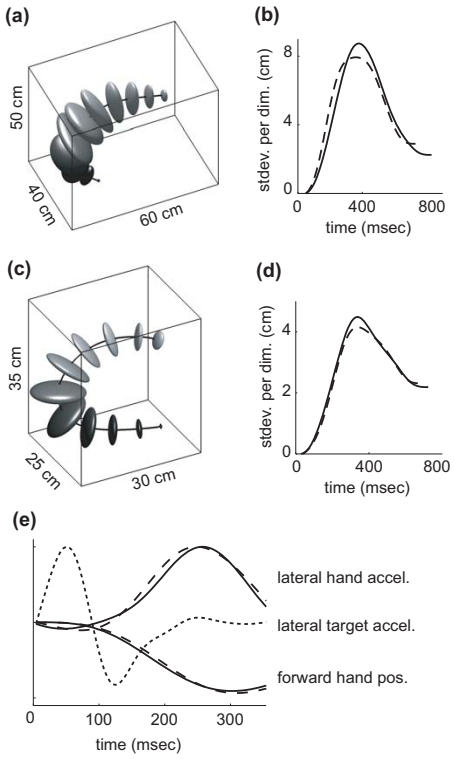


Figure 3.5 Variability during movement in Experiment 2-3. **(a)**, Spatial variability of unperturbed hand paths in Experiment 2. The ellipsoids correspond to ± 2 standard deviations in each direction. Aligning 3D trajectories for the purpose of computing variance is nontrivial, and was done as follows. We first resampled all movements for a given subject at 100 points equally spaced along the path, and found the average trajectory. Then, for each point along the average trajectory, we found the nearest sample point from each individual trajectory. These nearest points were averaged to recompute the corresponding point along the average trajectory, and the procedure was repeated until convergence (which only takes 2-3 iterations). In this way we extracted the spatial variability of the hand paths, independent of timing fluctuations. That is why the covariance ellipsoids are flat in the movement direction. **(b)**, Variability per dimension, for the stop (solid) and hit (dashed) conditions in experiment 2. At each point along the path, this quantity was computed as the square root of the trace of the covariance matrix for the corresponding ellipsoid, divided by 3. To plot variability as a function of time, we resampled back from equal-space to equal-time intervals. **(c,d)**, Same as subplots (a,b) but for experiment 3. **(e)**, Normalized target acceleration in the lateral direction, lateral hand position, and hand position in the forward direction (positive is towards the robot). Dashed lines – hit condition; solid lines – stop condition. Note that the onset of hand acceleration occurs before the movement reversal in the forward direction.

3.3.6 Flexible strategies for opportunistic control

The sensorimotor strategies predicted by optimal feedback control exhibit great flexibility – in the sense that they are adapted to the task, body and environment, and take advantage of every opportunity for achieving higher performance. This is in sharp contrast with traditional trajectory-planning models [10, 11] which essentially view all tasks as being the same as long as the average trajectory is the same. Here we make four additional experimental observations illustrating the flexibility inherent in the optimal control framework.

First, in the hit condition in experiments 2 and 3 subjects actually hit the target harder, even though they were not instructed to do so (and half of them had already performed the task in the stop condition). In both experiments the forward hand velocity before impact was 2-3 times larger in the hit condition compared to the stop condition. In experiment 3, where we used a force sensor, the normal force in the first 50msec after contact was about 3 times larger in the hit condition compared to the stop condition. Both the model and the experimental results suggest that the stopping requirement causes decreased sensitivity to positional errors late in the movement. Thus the relaxed stopping requirement in the hit condition is exploited to increase positional accuracy. It should be noted however that the stopping requirement was not fully eliminated in the hit condition. In experiment 2 (unperturbed trials, hit condition) subjects reduced their hand speed from 175cm/s peak to 39cm/s before contact – a 78% reduction; in experiment 3 this reduction was 84%. For comparison, the speed reduction in the stop condition was 96% and 92% respectively.

Second, subjects exploited the relaxed accuracy requirement in the vertical direction in experiment 3 – where the target was a vertical stripe rather than a circle. Focusing on unperturbed trials, we see that lateral and vertical endpoint errors are equally variable in experiment 2, but vertical "errors" are significantly more variable than lateral errors in experiment 3 (Fig. 4.6(a)). Furthermore, in experiment 3 subjects did not fully utilize feedback in order to adjust their vertical hand position relative to the falling target. Indeed, variability of the vertical endpoint position in absolute coordinates (relative to the room) was smaller than variability relative to the target. We know that subjects are able to correct in the lateral direction – not completely, but still the undershoot is much smaller than the correction. Thus the difference between absolute and relative variance in the vertical direction does not reflect an inability to correct, but rather an absence of a need to correct – in agreement with the minimal intervention principle [4].

Third, subjects found a way to exploit the different methods we used to detect the movement end in experiments 2 and 3. Here we focus on late perturbation trials and analyze the hand velocity immediately before contact with the target. In experiment 2 we used a speed threshold which required both forward and lateral velocity to be reduced in order to end the trial. In the hit condition the necessary velocity reduction could result from contact with the target, due to target impedance and friction respectively. In the stop condition, however, the target could not be exploited to stop the movement in either direction – and thus the lateral velocity in experiment 2 was small (Fig. 4.6(b)). In experiment 3 we used a force sensor to detect contact with the target, and defined the time of contact as the movement end, so lateral velocity did not have to be reduced as much. Subjects took advantage of this: the difference in lateral velocity between the stop and hit conditions in experiment 3 was much smaller than in experiment 2, and was not significant (Fig. 4.6(b)). We already know that the stopping requirement conflicts with positional accuracy. It is then likely that finding a way to partially avoid this requirement (in the lateral direction) afforded improved positional accuracy in experiment 3.

Fourth, subjects exploited the biomechanical redundancy of the arm when they had a chance. In experiments 2 and 3 redundancy was reduced by bracing the wrist; however we performed an earlier pilot experiment where the wrist was not braced (otherwise it was similar to experiment 2). In that case the lateral correction was accomplished with a combination of wrist flexion/extension and humeral rotation. We found that the percent contribution of the wrist was larger in late perturbations compared to early perturbations, in both the hit and stop conditions (Fig. 4.6(c)). The fact that the wrist contributes less than 30% in early corrections suggests that the preferred strategy is to use humeral rotation. This may be because for a given force level larger muscles are less affected by signal-dependent noise [27]. In late corrections, however, it is perhaps more difficult to accelerate and deccelerate the entire forearm within the remaining time, and thus the wrist contribution increases.

3.3.7 Modeling changes in duration and variability

The LQG framework, which we used in the above model as well as in most of our prior work on optimal feedback control, is computationally efficient but has a number of limitations. In the present context its limitations are: (i) movement duration cannot be modified online in response to target perturbations; (ii) stopping constraints have to be modeled with quadratic

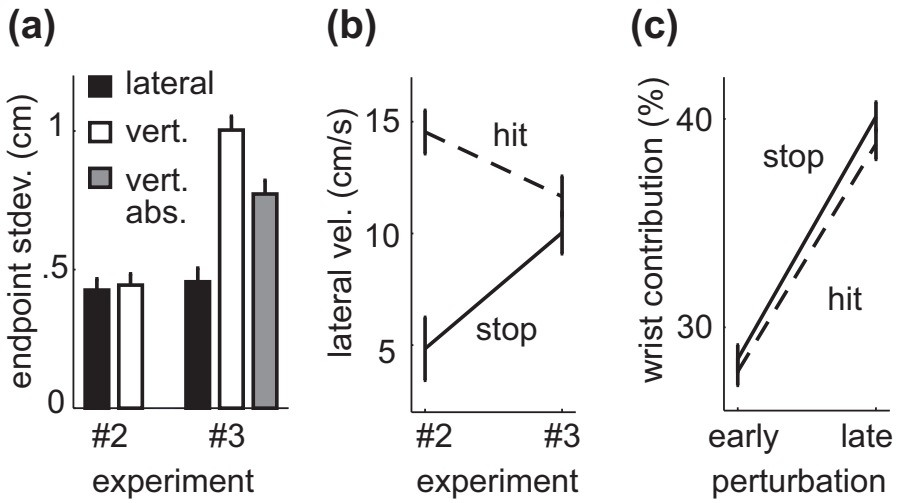


Figure 3.6 Endpoint standard deviation, lateral velocity and writst contribution in Experiment 2-3. **(a)**, Endpoint standard deviation in different directions, experiments 2 and 3, unperturbed trials. Black – lateral direction; white – vertical direction (coordinates relative to the target); gray – vertical direction (absolute coordinates). In experiment 3 the relative and absolute endpoint positions are different in the vertical direction, because the target is falling and the variability in movement duration causes variability in vertical target position at the end of the movement. **(b)**, Lateral velocity immediately before contact with the robot, in late perturbation trials. **(c)**, Wrist contribution to the lateral correction, in a pilot experiment with ten subjects. The main difference from experiment 2 was that the wrist was not braced. The lateral correction could be accomplished with humeral rotation (resulting mostly in translation of the hand-held pointer) or wrist flexion/extension (resulting in rotation of the pointer in the horizontal plane). The pointer was held in such a way that the Polhemus sensor was near the wrist. Therefore the lateral displacement of the sensor on perturbed trials (relative to the average trajectory on unperturbed trials) can be used as an index of how much humeral rotation contributes to the correction. The displacement of the tip of the pointer is defined as the total correction. The difference between the two is the contribution of the wrist. Dividing the latter by the total correction, and multiplying by 100, we obtain the percent wrist contribution.

costs instead of more natural step-function costs; (iii) the controller cannot be adapted to the statistics of the target perturbations. Here we present an optimal feedback control model which avoids these limitations. The new model is constructed using more general but less efficient discretization techniques which require a simpler 2nd-order model of arm dynamics. Movement duration is defined as the point in time when the hand first reaches the target plane. A term proportional to movement duration is included in the cost function. Time-out errors and hitting-hard errors are penalized with step-function costs. The tri-modal distribution of final target positions is taken into account in the optimization process. See Methods for details.

The new model (Fig. 3.7) accounts for the salient findings in Experiment 2 (Fig. 4.4(c)(d)(e)) and Experiment 3 (Fig. 4.4(f)(g)(h)). The speed of the corrective movement (slope of the positional traces in **Fig 7a** and Fig. 4.4(c)(f)) is smaller in early perturbations. The undershoot in late perturbations is larger in the stop versus hit condition. The undershoot in early perturbations is smaller compared to late perturbations, in both stop and hit conditions. Movement duration in baseline and early perturbations is larger in the stop versus hit condition. For late perturbations movement duration increases in both the stop and hit conditions. Note that the changes in movement duration are now predicted by the model, as opposed to being taken from the data as in the LQG model. This is possible because the new controller can adjust the duration online – by modulating the speed in the main movement direction and thus reaching the target plane at different times.

The new model allows us to address an additional phenomenon which is beyond the scope of LQG models. The phenomenon (Fig. 3.8) is that trajectory variability on unperturbed trials was larger in experimental sessions with perturbations compared to baseline sessions without perturbations. Results from the hit and stop conditions are averaged in this analysis. **Fig 8** shows that frequent perturbations lead to some adaptive change in the sensorimotor system which in turn leads to increased variability on trials without perturbations. The precise time-course of such adaptation is difficult to estimate because measuring variability requires many trials.

What could be the nature of this adaptive change? One possible explanation is that, in sessions with perturbations, trial-to-trial adaptation [31, 32] causes the system to be in a different state every time an unperturbed trial is encountered. However in target perturbation paradigms trial-to-trial adaptation is negligible [77], see also below. Another possible explanation is that

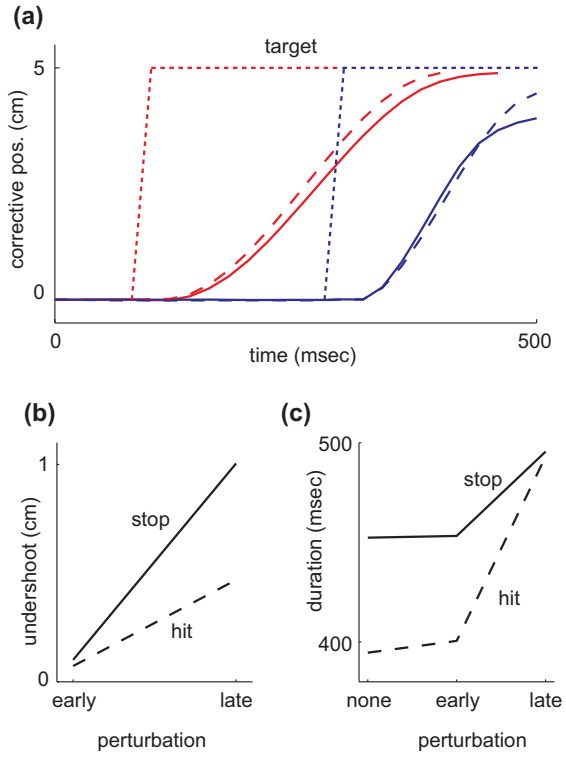


Figure 3.7 A MDP model for Experiment 2-3. **(a)**, Corrective movements of the more general optimal feedback control model. The solid and dashed lines correspond to the stop and hit conditions respectively. The hand is restricted to a grid of discrete states, however the dynamics are stochastic, and so the average (over 1000 simulated trials) is smooth even though the individual trajectories have a staircase pattern. **(b,c)**, Undershoot and movement duration in the stop and hit conditions for different perturbation times. Same format as the experimental data in Fig. 4.4.

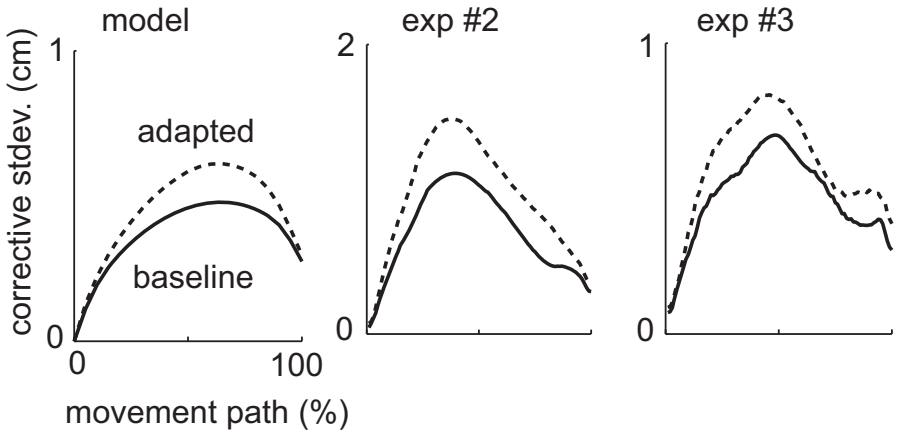


Figure 3.8 Hand positional variance on unperturbed trials from MDP model and Experiment 2-3. Hand positional variance on unperturbed trials, measured along the perturbation direction. Trajectories are aligned at equal intervals along the movement path in order to compute variance. Solid line (baseline) is the variance in blocks without perturbations. Dashed line (adapted) is the variance in blocks with 66% perturbations. Data from the hit and stop conditions are averaged.

target perturbations are for some reason misinterpreted as an increase in sensory noise – in which case the "optimal" thing to do is reduce the reliance on sensory feedback, causing suboptimal performance. A third explanation, which we pursue below, is an adaptive change in the feedback controller.

In environments with large unpredictable perturbations one would expect the optimal feedback controller to be more concerned with correcting the perturbations than the smaller errors due to internal noise. This is confirmed by our simulations. In Fig. 3.8(a) we compare the trajectory variability of two feedback controllers, one optimized for an unperturbed environment and the other one for a perturbed environment matching our task. As expected the latter controller is better at correcting for perturbations (data not shown), however it allows higher variability on trials without perturbations. This is broadly consistent with the minimal intervention principle as well as with the idea of Pareto optimality: improving any aspect of the behavior of an optimal controller requires sacrifices elsewhere. Analyzing the specific changes in feedback gains which lead to increased variability is interesting but beyond the scope of the present paper.

3.3.8 Lack of trial-to-trial adaptation

Trial-to-trial adaptation has been found in perturbations of the hand [31, 32] but not the target [77]. Here we replicate the latter finding: in our target perturbation experiments trial-to-trial adaptation turns out to be negligible.

To quantify such adaptation we adopted a version of the state-space approach, namely

$$\mathbf{z}(n+1) = A\mathbf{z}(n) + B\mathbf{w}(n) + \eta(n)$$
$$\mathbf{y}(n) = C\mathbf{z}(n) + D\mathbf{w}(n) + \gamma(n)$$

The correction on trial n is denoted $\mathbf{y}(n)$ and has two elements: distance in the corrective direction from the average unperturbed trajectory, measured at the time of the late perturbation and at the end of the movement. The perturbation $\mathbf{w}(n)$ has two elements specifying the target position (again in the corrective direction) immediately before the late perturbation and at the end of the movement. The vector $\mathbf{z}(n)$ is the internal learning state. It also has two elements which the model is free to use in whatever way is needed to fit the data. We quantify the correction and the perturbation using pairs of measurements so as to allow the model the capture the difference between early and late perturbations. $\eta(n), \gamma(n)$ are independent zero-mean 2D Gaussian random variables with covariances Q, S.

The next learning state $\mathbf{z}(n+1)$ may in general depend on the current learning state $\mathbf{z}(n)$, the perturbation $\mathbf{w}(n)$ and the correction $\mathbf{y}(n)$. However $\mathbf{y}$ is a linear function of $\mathbf{z}, \mathbf{w}$ and so we do not include a $\mathbf{y}$-term in the first equation. Note also that if there is any trial-to-trial learning here it should be related to predicting the final target position (and initiating a normal movement aimed at that position); thus it makes more sense to learn from $\mathbf{w}$ rather than $\mathbf{y}$.

The sequences of corrections $\mathbf{y}(n)$ and perturbations $\mathbf{w}(n)$ were measured. Given these measurements, the most likely values of A, B, C, D, Q, S as well as the sequence of learning states $\mathbf{z}(n)$ were computed using the EM algorithm [76]. Since EM can get trapped in local mimima, model fitting was run multiple times with different initial conditions and the best result was used. We fit the model separately for each experiment (2 vs. 3), condition (stop vs. hit) and subject. The first 2/3 of the data in each experimental session were used for model fitting and the last 1/3 for model testing.

To evaluate the performance of the learning model we regressed each component of $\mathbf{y}$ either on $\mathbf{z}$ and $\mathbf{w}$, or on $\mathbf{w}$ alone, or on $\mathbf{z}$ alone. The first regression measures the performance of

Table 3.2 $100R^2$ for regressions on fitted model

variable	regressor	all trials	baseline	early jump	late jump
y_1	$\mathbf{z}, \mathbf{w}$	74.36*	·	91.13*	·
y_1	$\mathbf{w}$	71.50*	·	86.33*	·
y_1	$\mathbf{z}$	·	·	·	·
y_2	$\mathbf{z}, \mathbf{w}$	99.34*	·	99.85*	98.15*
y_2	$\mathbf{w}$	99.28*	·	99.83*	97.34*
y_2	$\mathbf{z}$	·	·	·	·

the full model while the latter two regressions measure the contribution of feedback correction ($\mathbf{w}$) and learning component ($\mathbf{z}$) respectively. The R^2 and p values for the regressions were averaged over subjects, experiments and conditions. The regressions were frist done on all trials, and then separately on the baseline (no perturbation) trials, early jump trials and late jump trials – because the models are likely to perform differently on different trial types. Table 1 shows the average R^2 values multiplied by 100 (to obtain a measure of variance explained), only for the cases where $p < 0.05$ on average. In the remaining cases we found $p > 0.3$, thus there was clear separation between significant and non-significant regression fits.

None of the regressions on $\mathbf{z}$ alone are significant, suggesting lack of trial-to-trial learning. The regressions on $\mathbf{w}$ are as expected: given the target position at the middle and at the end of the movement, one can predict the hand position at the end of the movement (y_2) in both types of perturbations as well as the hand position at the middle of the movement (y_1) in early perturbations. Note that the combined model $\mathbf{z}, \mathbf{w}$ is slightly but systematically better than $\mathbf{w}$ alone, suggesting that there may be a small learning effect. Given how small this effect is, it is not surprising that regressions on $\mathbf{z}$ alone were far from significant.

3.4 Discussion

The biological processes that continuously improve behavior closely resemble iterative optimization. This makes optimal control theory a natural framework for studying the neural control of movement. It is also a very successful framework in terms of explaining the details of experimental data [17]. However one of its most appealing features remains largely untapped: the ability to predict task-specific sensorimotor strategies and thereby changes in behavior that

result from systematic task variation. This is a gap not only in optimal control models but in the field of motor control in general. A substantial number of studies (including most of the literature on motor adaptation) have used a single task: reaching. The emphasis on servo control has created the impression that as long as a desired trajectory can somehow be planned, motor execution (and all sensorimotor processing during movement) is the same no matter what the organism is trying to accomplish. Planning models have focused on the geometry of limb trajectories and have largely ignored the context which gives functional meaning to these trajectories. We see this as a substantial gap in the current understanding of sensorimotor function; the present paper is a step towards filling that gap.

In order to vary the task systematically, we need a compact and experimentally accessible representation of the task space. Optimal control provides the perfect tool: composite cost functions. A central argument of this paper is that subjects optimize a composite cost as opposed to a homogeneous cost under multiple hard constraints. Indeed we did everything we could to enforce a hard constraint on movement duration, and yet subjects never treated it as such. Instead they always found a balance between undershoot and time-out errors. The changes in our experimental design affected the relative importance of these errors; in particular the switch to the intercept task (which made the duration threshold explicit) resulted in the lowest percent time-out errors. In addition to accuracy and duration, we proposed that the composite cost includes endpoint stability (stopping in particular) and energy consumption. We showed directly that stability is part of the cost – by allowing subjects to interact with a high-impedance target and finding that they take advantage of it. The only evidence for the energy cost was indirect: it was needed to make our model fit the data. However other studies have provided more direct evidence: increased muscle cocontraction has been found to yield more accurate movements [83, 105], and yet this is not a strategy that subjects normally use – suggesting that they care about energetic efficiency in addition to accuracy. It is also notable that successful optimal control models of full-body movements are predominantly based on energy minimization [23, 106]. In obstacle avoidance tasks (experiments 2 and 3) there is likely to be a fifth component of the cost having to do with avoiding the obstacle. Although we did not model this cost, we showed that obstacle avoidance does not rely on hard constraints such as fixed imaginary targets to the side of the obstacle.

The main effect we analyzed – the incomplete correction for late perturbations – re-

flects the closed-loop component of the sensorimotor strategy. The fact that the effect decreased in the hit condition means that the visuo-motor loop operated differently, as predicted by our model. Thus changes in stopping requirements (as well as target impedance) caused changes in the way visual feedback is used to make online corrections. This may be the first demonstration that visuo-motor feedback loops are affected by the task and in particular by non-visual components of the task. In addition to demonstrating task sensitivity, we provided further evidence that sensorimotor strategies are consistent with the minimal intervention principle of optimal feedback control [4]. We found that positional variability is large during movement (especially 3D movement) and is only reduced near the end, where accuracy is needed. We also found that when the target is a vertical stripe, endpoint variability is larger in the vertical direction and visual feedback is not fully utilized to suppress variability in that direction. These results reaffirm the utility of looking beyond average trajectories, studying variability patterns and responses to perturbations, and modeling the sensorimotor strategies responsible for such effects.

Motor adaptation is a phenomenon which has not yet been addressed in the optimal control framework, but in principle is easy to address, as we showed in our model of increased variability due to frequent perturbations. One can impose any change in the task or environment, compute the new optimal controller and use it as a model of adapted behavior. Of course adaptation is rarely complete, thus the predicted adaptation effect should be somewhere in between the baseline and fully-adapted optimal controllers. An interesting open question is how to relate trial-to-trial dynamics of learning to asymptotic predictions regarding optimal adaptation. One way to do this is to model trial-to-trial changes as arising from an iterative optimization algorithm which in the limit converges to the adapted optimal controller. This approach may yield richer models of learning dynamics than the linear state-space models currently used.

Acknowledgment

This chapter, in part, was originally published in *the Journal of Neuroscience*, 27(35): 9354-9368,2007. The dissertation author was the primary author of the paper.

Chapter 4

Smart Learning

4.1 Introduction

Everyday experience suggests that we are able to learn from a changing environment and adjust our movements accordingly. This remarkable adaptability of the sensorimotor system has been demonstrated in adaptation experiments using visual perturbations [28, 29, 30] as well as force-fields produced by a robotic manipulandum [8, 31, 32]. The common finding is that an internal model of the experimentally imposed perturbation is acquired, and then used to generate predictive compensation [33, 34]. Despite extensive work on the statistical formulation of internal models [35, 30, 36, 37, 38] the question of how the CNS uses these internal models to produce motor commands has received surprisingly little attention. There is an implicit assumption in the literature that the output of internal models is in one-to-one correspondence with changes in motor behavior, or in other words the control system itself is not changing. This assumption has no reason to be true and indeed we show here that it can be systematically violated. In particular, we show that the same perturbation sequence can elicit very different forms of adaptation as far as motor behavior is concerned, even though the information content (and presumably the internal model being formed) is the same. We account for these differences in the framework of optimal feedback control.

The evidence for flexible control strategies during both motor planning and motor execution has been demonstrated in previous work [4, 71, 16]. However, these optimal feedback control models are essentially a description of the biological system in its steady state. They

have not addressed how the control system responds when the motor plant undergoes changes, especially the flexible adjustments based on knowledge about the statistical properties of the environment learned from past experience. This limitation is partly due to the fact that optimal feedback control is in general hard to solve, except for the case of linear dynamics, quadratic costs and Gaussian noise (LQG) [66]. Although a recent study successfully applied the LQG framework to explaining the curvature in saccades after adapting to target perturbations, it did not model the statistical distribution of the perturbations explicitly [68]. Here we applied an iterative LQG algorithm [67] which enables us to handle non-Gaussian perturbation statistics and non-quadratic cost functions necessary to model our data.

Our point of departure is the puzzle why strong trial-by-trial learning [31, 32] has been observed in response to visual and mechanical perturbations but not to target jumps [32, 77, 71]. We first designed a reaching experiment where the target was perturbed in a non-random way at different times during the movement and in different dynamic environments. We found that subjects modified their behavior based on the statistics of the perturbations, but only in conditions where lack of adaptation would have been detrimental to task achievement. In particular, no learning effect was observed when the target perturbation occurred early in the movement and subjects had enough time to make an on-line correction. To understand whether this lack of predictive responses in early perturbations was attributable to a lack of learning, we designed a second experiment where we tested subjects both before and after forcing them to predict. We found that subjects preferred not to rely on prediction in early-perturbation trials even after learning to do so in a different condition. One possible reason is that, even though the perturbation direction was predictable in a statistical sense, the target could still jump in a different direction and thereby require large on-line correction. If the motor system is sensitive to such outliers, limiting the perturbation to one direction should increase the predictive response. This hypothesis was tested in a third experiment where perturbations were generated from one of the two distributions with the same mean but different shapes: both left and right perturbations were possible in one case but only right perturbations were possible in the other. We found that when outliers were eliminated the predictive response increased. This was modeled by introducing an energy cost which increases faster than the usual quadratic cost.

The flexible use of internal models in generating motor commands, which we observed in all three experiments, could be largely explained within the framework of optimal feedback

control. This required some novel extensions to the framework. The extensions have to do with robustness (as in high-order energy costs) as well as a preference for control strategies that tend to be successful in everyday life outside the lab.

4.2 Materials and Methods

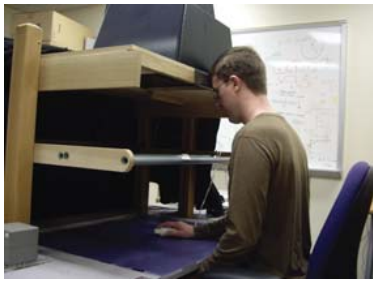
4.2.1 Experimental Design

Experimental Setup

In all three experiments, we used a similar setup. Subjects made planar reaching movements on a table positioned at chest level (Fig. 4.1a). A 21 inch flatscreen monitor was mounted above the table facing down and was viewed in a see-through horizontal mirror. In this way computer-generated images could be physically aligned with the hand workspace. Subjects held in their right hand a small pointer with an electromagnetic Polhemus (Colchester, VT) Liberty sensor and a 3D accelerometer attached to it. The Polhemus sensor measured three-dimensional position at 240Hz, and the accelerometer measured the acceleration at 960Hz. The room was dark to prevent subjects from seeing their hands. Visual feedback of hand position was provided by a 2-cm-diameter red disk on the computer monitor, referred to as the hand cursor. A 2-cm-diameter green disk represented the target. The task was to move the hand cursor to a starting position, wait for the target to appear, and move to the target after a ready signal was released. The main movement was on the vertical direction from bottom to up. During the movement, the target was either stationary or rapidly displaced left or right relative to the subject. Subjects were instructed that perturbations may occur and asked to always move to the final target position within the allowed time limit.

Movement onset was detected on-line using a 0.1cm threshold on the distance between the current hand position and the hand position when the ready signal was released. Movement was terminated when target was hit within the allowed time interval (successful trial), when a horizontal boundary line going through the target was crossed without hitting the target (missing error), or when the allowed time exceeded (time-out error).

A



B

	RANDOM	EARLY	RANDOM	LATE	RANDOM	COG
trials	20	80	20	80	20	80
perturbation time (ms)	175	0	175	350	175	N/A

C

	testing	training		testing
	EARLY	EARLY with COG	EARLY with COG	EARLY
trials	40	80	80	40

D

	left	baseline	right
bimodal	0	50	50
tri-modal	20	10	70

Figure 4.1 Experimental setup and paradigm. **(a)**, Experimental setup. **(b)**, Chronology of Experiment 1. **(c)**, Chronology of Experiment 2. **(d)**, Chronology of Experiment 3.

Experiment 1

After a brief familiarization session, twelve subjects performed 20 baseline trials without perturbations, followed by six blocks of trials with perturbations (Fig. 4.1b). The main blocks were EARLY, LATE and COG, each separated by a RANDOM block to eliminate the interference from the previous block. The three main blocks had the identical sequence of perturbations applied to the target. Each sequence contained 10 baseline trials, 35 left perturbation trials, and 35 right perturbation trials. Perturbation trials were not permuted randomly, but ordered such that once a perturbation occurred, it would repeat to the same direction in the next 2-4 trials in a row with different probabilities (Table 1). The sequence was generated using a first order Markov Model with added constraints on the number of trials with repeating perturbations, leading to this uneven distribution. Once a baseline trial occurred, it was followed by either a left or a right perturbation with equal probabilities. Despite the same perturbation sequence, perturbations times were different in these three blocks. The target was visible all the time in EARLY and LATE but was perturbed at onset of the movement in EARLY and 350ms into the movement in LATE. In COG, in contrast, the target was invisible during the movement and subjects were asked to move to the predicted final target position, with visual feedback of the true final target position provided briefly after the movement was finished (the horizontal boundary line was reached).

In addition to varying perturbation times, we also changed the dynamics of the hand cursor in either the DIRECT condition or the SPRING condition. In the DIRECT condition, the hand cursor reflected the true hand position. In the SPRING condition, in contrast, the hand cursor was simulated as a point mass attached to the hand via a spring and thus had its own dynamics. Parameters of this mass-spring-damper system were set as follows: mass $m = 1$kg, spring constant $\alpha = 100$N/m, and viscous damper of damping coefficient $\beta = 20$Ns/m. As a result, the hand cursor lagged behind the real hand and could not change directions as rapidly. Since the parameters satisfied the critical damping criterion $\beta^{\mathbf{2}} = 4m\alpha$, stability of the hand cursor was guaranteed.

Subjects were divided into two groups, and were tested in the six blocks under either the DIRECT condition (group 1) or the SPRING condition (group 2). Within each group, one half of the subjects performed the EARLY block first, and then the LATE block; this order was reversed in the other half; COG was always the final block. In all these blocks/conditions,

Table 4.1 Probability of perturbation directions after perturbing the target to a different direction from the previous trial

trial number after first perturbation	same direction	opposite direction	baseline
3	53	29	18
4	89	11	0
5	0	63	37

subjects were required to finish the movement within 800ms.

Experiment 2

Subjects were tested in two identical blocks before and after they were forced to learn to predict the perturbations in a training block (Fig. 4.1c). In both training and testing blocks, the perturbation sequence followed the same pattern as in Experiment 1. In the testing block, target was visible all the time and all perturbations occurred at the onset of the movement. In the training block, however, to force subjects to learn to predict the perturbations, in 1/8 of the trials the target became invisible before the movement and subjects were asked to move to the predicted target position. The true final target position was provided briefly after the movement was finished. Subjects could not predict when these cognitive trials occurred, and had to predict the perturbation in the following trial after each movement. Each testing block had 40 trials, long enough to get a stable learning effect based on results from Experiment 1. The training block had 160 trials to make sure the perturbation pattern was learned.

Sixteen subjects participated in the experiment. The target was visible all the time in the experiment except for the cognitive trials, and the hand cursor reflected the true hand position.

Experiment 3

Different from the repeat-and-switch pattern in Experiment 1 and 2, here perturbations were independently generated from either a bimodal distribution in the bimodal block, or a tri-modal distribution in the tri-modal block. Let $\mathbf{j}$ be a vector representing the probability of left perturbation, baseline, and right perturbation, then $\mathbf{j}_{\text{bimodal}} = [0.0, 0.5, 0.5]$ and

$\mathbf{j}_{\mathrm{tri-modal}} = [0.2, 0.1, 0.7]$. These two distributions were chosen because they had the same mean but different shapes, especially, outliers only occurred in the latter where the perturbation could change directions from left to right, or vice versa. In perturbation trials, the target was rapidly displaced 10cm left or right at 150ms during the movement. In both blocks, subjects were required to finish the movement within 700ms. The target was visible all the time in the experiment, and the hand cursor reflected the true hand position.

After a familiarization session, Sixteen subjects performed both the bimodal block and the tri-modal block, with 100 trials in each (Fig. 4.1(d)). Before starting each block, subjects were instructed whether the target would be displaced to both left and right, or only right. One half of the subjects performed the bimodal block first, and then the tri-modal block; this order was reversed in the other half.

4.2.2 Statistical analysis

To get a stable learning effect in response to target perturbations in Experiment 1-3, only data from the second half of each block were used (see Fig. 4.2 and Fig. 4.5). In the comparison of hand movements on the lateral direction (corrective direction) in Experiment 1 and 2 (see Fig. 4.2(b)(c)(d)(e)), only adaptation trials, defined as trials where target was perturbed to the same direction as in the previous trial, were used. To analyze the learning process, all adaptation trials were included in Experiment 1 (see Fig. 4.6(a), 4.6(b)), and all data in the bimodal block were used in Experiment 3 (see Fig. 4.6(c)).

Statistical tests were based on n-factor ANOVA ("anovan" in the MATLAB Statistics Toolbox). We avoided averaging to the extent possible. In the comparisons of the lateral hand movement prior to perturbations (see Fig. 4.2(b), Fig. 4.5(b)), earliest correction time (see Fig. 4.2(c)), and the endpoint lateral error (see Fig. 4.2(d)) to study the learning effect in Experiment 1-3, individual trials were treated as repeated measures, and the factors were the experimental conditions (EARLY vs. LATE vs. COG, DIRECT vs. SPRING in Experiment 1; bimodal vs. tri-modal in Experiment 3) and perturbation directions if applicable (see Fig. 4.2(c)). In the comparisons of lateral hand positions before the late perturbation time in the SPRING condition in Experiment 1 (see Fig. 4.6(a)), endpoint lateral error and the corresponding correction prior to perturbations through the LATE block in Experiment 1 (see Fig. 4.6(b)) and through the bimodal block in Experiment 3 (see Fig. 4.6(c)) to study the learning process, individual

trials were treated as repeated measures, and the factors were trial numbers and experimental conditions if applicable (EARLY vs. LATE vs. COG in Fig. 4.6(a) DIRECT vs. SPRING in Fig. 4.6(b)). All comparisons of means were based on Tukey's criterion for post hoc hypothesis testing. Differences were reported as significant when $p < 0.05$. The error bars shown in the figures correspond to 1 pooled SE of the mean, as computed by the multiplecomparison function used to perform Tukey's test ("multcompare" in the MATLAB Statistics Toolbox). In the comparison of corrections in response to right perturbations in the two blocks in experiment 3 (see Fig. 4.5(a)), all right-perturbed trials that a subject performed in each block were combined to obtain a single number, and then a t-test was used and the only factor was the block.

4.2.3 Optimal feedback control model (ILQG)

We model the hand as a $m = 1$kg point mass moving in a horizontal plane, with viscosity $b = 10$Ns/m approximating intrinsic muscle damping. The point mass is driven by two orthogonal force actuators which can both pull and push (approximating two pairs of agonist-antagonist muscles). The actuators act as muscle-like first-order low-pass filters of the control signals, with time constant $\tau = 0.05$s. These settings of m, b, τ are chosen to be compatible with biomechanics and are not adjusted to fit the data. For the mass-spring-damper system in the SPRING condition in experiment 1, spring constant $\alpha = 100$N/m and viscous damper of damping coefficient $\beta = 20$Ns/m are kept the same as in the experiment.

Let $\mathbf{p}(t), \mathbf{v}(t), \mathbf{a}(t), \mathbf{u}(t), \widetilde{\mathbf{p}}(t), \widetilde{\mathbf{v}}(t), \mathbf{p}^*(t)$ be the 2D hand position, hand velocity, hand actuator state, control signal, cursor position, cursor velocity, and target position respectively. The corresponding units are cm, cm/s, N, N, cm, cm/s and cm. Time axis was discretized with time step $\Delta t = 20$ms. Time index is $t \in [0, t_f]$, where the final time t_f is specified according to the experimental data. The hand dynamics are modeled as

$$\begin{aligned}
\mathbf{p}(t+1) &= \mathbf{p}(t) + \mathbf{v}(t)\,\Delta t \\
\mathbf{v}(t+1) &= \mathbf{v}(t) + \frac{1}{m}\left(\mathbf{a}(t) - b\mathbf{v}(t)\right)\Delta t \\
\mathbf{a}(t+1) &= \mathbf{a}(t) + \frac{1}{\tau}\left(\mathbf{u}(t) - \mathbf{a}(t)\right)\Delta t
\end{aligned}$$

and the hand cursor dynamics are modeled as

DIRECT	SPRING
$\widetilde{\mathbf{p}}(t) = \mathbf{p}(t)$	$\widetilde{\mathbf{p}}(t+1) = \widetilde{\mathbf{p}}(t) + \widetilde{\mathbf{v}}(t)\,\Delta t$
$\widetilde{\mathbf{v}}(t) = \mathbf{v}(t)$	$\widetilde{\mathbf{v}}(t+1) = \widetilde{\mathbf{v}}(t) + \frac{1}{m}(\alpha(\mathbf{p}(t) - \widetilde{\mathbf{p}}(t)) - \beta\widetilde{\mathbf{v}}(t))\Delta t$

In order to define an optimal control problem we also need a cost function. As in our previous work [71], we use a mixed cost function defined as

$$\|\mathbf{p}^* - \tilde{\mathbf{p}}(t_f)\|^2 + w_{\text{stop}}\|\tilde{\mathbf{v}}(t_f)\|^2 + w_{\text{energy}} \sum_{t_i=0}^{t_i=t_f-1} \|\mathbf{u}(t_i)\|^4 \Delta t$$

The three cost terms encourage endpoint positional accuracy, stopping at the target, and energetic efficiency, respectively. The weights w_{energy} and w_{stop} are unknown parameters. Note the energy cost is not quadratic but 4^{th} order. This will be explained later.

The initial state is $\mathbf{p}(0) = \mathbf{v}(0) = \widetilde{\mathbf{p}}(0) = \widetilde{\mathbf{v}}(0) = [0; 0]$. The default target position $\mathbf{p}^*$ is $[0; y^*]$ but can be perturbed to either $[x^*; y^*]$ or $[-x^*; y^*]$. Consistent with the experiment, we set $x^* = 8\text{cm}$, $y^* = 20\text{cm}$ for Experiment 1; $x^* = 10\text{cm}$, $y^* = 18\text{cm}$ for Experiment 3. Again, let $\mathbf{j}$ be the vector representing the probability of left perturbation, baseline and right perturbation. We set $\mathbf{j}_{\text{bimodal}} = [0.0, 0.5, 0.5]$ and $\mathbf{j}_{\text{tri-modal}} = [0.2, 0.1, 0.7]$ using values from Experiment 3. For Experiment 1, if we ignore the constraints on the number of sequence of trials with the same perturbations and simply compute the probability of perturbing the target to the three directions after a perturbation trial, we get $\mathbf{j} = [0.1, 0.1, 0.8]$ after flipping the left perturbations to the right, as used in the model. We do not model explicitly the visuomotor delays or uncertainty in detecting the target observations. To obtain correct reaction times, we simply model each target perturbation as occurring 120ms later than the corresponding experimental perturbation.

However, this optimal control problem does not fit in the well-developed Linear-Quadratic-Gaussian (LQG) formalism due to two reasons. First, $\mathbf{p}^*(1)$ (lateral endpoint target position) is not Gaussian but rather a mixture of 3 delta functions. Second, the energy cost is not quadratic, which will prove to be crucial in explaining data in Experiment 3. Another option is to discretize the state and control spaces and convert it into a Markov decision problem (MDP) [69] which can be solved via dynamic programming [70]. We did try this on the DIRECT condition in Experiment 1 and got similar results as reported later when the dynamics was modeled as a

2^{nd} order system using six states. However, when the dimension of state goes higher (necessary in modeling the SPRING condition in Experiment 1), such method becomes inapplicable due to the curse of dimensionality.

Since performance of our perturbation experiments is measured based on only hand trajectories, we used the Iterative LQG (ILQG) [67]. This method iteratively uses linearizations of the nonlinear dynamics around the accurate trajectory, and improves that trajectory by minimizing a quadratic approximation to the optimal cost-to-go function. As a result, it can deal with nonlinear system such as the human arm, with more complex models of cost. In addition, it allows us to embed statistical distributions of external perturbations in the framework after making the following modifications: given the Bellman equation [107], optimal cost-to-go function $v(\mathbf{x}_t)$ (i.e. the cost expected to accumulate from now until the end of the movement, assuming optimal behavior) has the following form

$$v(\mathbf{x}_t) = \min_{\mathbf{u_{x_t}}} \left(\|\mathbf{u_{x_t}}\|^4 + E(v(\mathbf{x}_{t+1})) \right)$$

Since uncertainty only occurs at the perturbation time, we have

$$t \neq t_{\text{jump}}, E(v(\mathbf{x}_{t+1})) = v(\mathbf{x}_{t+1})$$
$$t = t_{\text{jump}}, E(v(\mathbf{x}_{t+1})) = \mathbf{j}(\text{left})v^{\text{left}}(\mathbf{x}_{t+1}) + \mathbf{j}(\text{middle})v^{\text{middle}}(\mathbf{x}_{t+1}) + \mathbf{j}(\text{right})v^{\text{right}}(\mathbf{x}_{t+1})$$

for $t \in [0, t_f]$, where left, middle and right refers to left perturbation, baseline, and right perturbation respectively. Thus, in each iteration we can first solve the control problem for the left, middle and right paths after the perturbation separately, and combine them at the perturbation time according to the probability of left perturbation, baseline, and right perturbation. This result is then used to initialize the path before the perturbation and solve the problem backwards. During this process, an affine control law is designed for the system and applied forward in time to the system to obtain an updated open-loop control along the way. For the next iteration, the splitting point is modified by applying this new open-loop control law and the above process repeats. Such iteration continues until the new sequence of control law and the old one are sufficiently close. Since this is essentially a local method and can get trapped in local minima, it was run multiple times with different $\mathbf{u}_0$ and the best result, which yields the minimum cost, is used.

Table 4.2 Parameters of iterative-linear-quadratic-Gaussian model

parameters	**criteria**
$m = 1\text{kg},\ b = 10\text{Ns/m},\ \tau = 0.05\text{s}$	compatibility with biomechanics
$\alpha = 100\text{N/m}, \beta = 20\text{Ns/m}$	same as in Experiment 1 (SPRING)
$x^* = 8\text{cm},\ y^* = 20\text{cm},\ \mathbf{j} = [0.1, 0.1, 0.8]$	same as in Experiment 1
$t_f^{\text{EARLY}} = 27,\ t_f^{\text{LATE}} = 34,\ t_f^{\text{COG}} = 19$	from data in Experiment 1 (DIRECT)
$t_f^{\text{EARLY}} = 32,\ t_f^{\text{LATE}} = 34,\ t_f^{\text{COG}} = 31$	from data in Experiment 1 (SPRING)
$x^* = 10\text{cm},\ y^* = 18\text{cm}$	same as in Experiment 3
$\mathbf{j}_{\text{bimodal}} = [0.0, 0.5, 0.5],\ \mathbf{j}_{\text{tri-modal}} = [0.2, 0.1, 0.7]$	same as in Experiment 3
$t_f^{\text{bimodal}} = 29,\ t_f^{\text{tri-modal}} = 29$	from data in Experiment 3
$w_{\text{stop}} = 0.00005, w_{\text{energy}} = 1E-10$	overall **fit** to data

From the MDP model we developed in a pilot study, we found that the control dependant-noise had very little effect on the results. Therefore, here we simply put this noise term to 0. We also dropped the additive noise since it has no effect on the control law in the absence of control-dependent noise in the model [67]. The model parameters, and the criteria for choosing their values, can be summarized as follows:

4.3 Results

4.3.1 On-line corrections to perturbed target after learning

Experiment 1 was to test whether subjects could learn the statistical properties of perturbations, and use such knowledge in on-line corrections to compensate for the delay of sensory feedback. Subjects made forward reaching movement on a horizontal table, while the target was

displaced in an orthogonal direction (left or right relative to the subject) unexpectedly in perturbation trials. Since no learning effect was observed when the direction of perturbation applied to target was random from previous experiments [77, 71], in this experiment, perturbations were carefully controlled such that once a perturbation occurred, it would be more likely to repeat to the same direction in the next trial (see materials and methods for details). The same perturbation sequence was tested in three blocks under two conditions. The three blocks differed in the way the target perturbation was represented were provided. Target was visible all the time but disturbed early in the EARLY block and late in the LATE block. Perturbation time was carefully controlled such that subjects could have enough time to make full corrections based on the visual feedback of the perturbation in EARLY but not LATE. In COG, on the other hand, the target was invisible during the movement to force subject to learn to predict the final target position. Each subject performed these three blocks under either the DIRECTION condition or the SPRING condition, differing in the hand cursor dynamics. In the DIRECT condition, the hand cursor reflected the true hand position; whereas in the SPRING condition, the hand cursor was simulated as a mass attached to the hand via a spring. Parameters in the mass-spring-damper system were chosen so that the hand cursor lagged behind the true hand and thus made the online-correction of the hand cursor much harder especially in the LATE block.

In the following analysis, only data from the second half in each block were included to get a stable learning effect. The learning process through trials are presented in section 4. Fig. 4.2(a) shows the average hand/cursor paths in EARLY, LATE and COG blocks under DIRECT and SPRING conditions. In the rest of the analysis, trials following the left perturbations are mirrored around the vertical axis and pooled with the corresponding trials following the right perturbations. In this way, the previous perturbation is always to the right (trials following baselines are not included). We define adaptation trials as trials where the target was displaced to the same direction as in the previous trial (to the right), catch no jump trials as baselines, and catch jump trials as trials where the perturbation was to the opposite direction to the previous trial (to the left). When the target was visible during the movement as in EARLY and LATE, all movements exhibit smooth corrections, regardless of the perturbation time and the hand cursor dynamics. The fact that initial hand paths to different perturbation directions are not separable shows that subjects have not responded to the perturbations. In COG, in contrast, the lack of curvatures in hand paths indicates that no visual target was available in guiding on-line correc-

tions. Note the discrepancy between the real hand (thin lines) and the hand cursor (thick lines) in SPRING, especially the delay of the latter. As can be seen, although the perturbation sequence applied to the target was identical in all these blocks/conditions, hand movements on the lateral direction (corrective direction) were quite different.

The different corrective movements in different blocks/conditions are further quantified using both the lateral hand position measured at 100ms after the early and late perturbation time (Fig. 4.2(b)), and the earliest correction time, measured as the first point in time where accelerations in left-perturbed trials differed significantly from those in right-perturbed trials (Fig. 4.2(c)). In Fig. 4.2(c) 0 represents the main movement onset detected using a positional threshold, which is thus later than that based on accelerations. As we expected, when no target perturbation was available during the movement as in COG, corrective movements started from the beginning of the main movement and became significantly bigger than 0 even measured at the early perturbing time. A similar learning effect was also observed in LATE-SPRING, indicating the use of prediction in on-line corrections to compensate for delays of the target perturbation and the hand cursor. In LATE-DIRECT and EARLY (both DIRECT and SPRING), on the other hand, corrections were not significant and were initiated much later with respect to the perturbation time, indicating little or no learning effect. Fig. 4.2(d) shows the endpoint lateral error, with COG being the largest, followed by LATE, and that in EARLY was close to 0. Note the similar accuracy in LATE with or without delaying the hand cursor, even the task was much more difficult in the former. The reason will be explained in section 4.

In addition to the average movement, Fig. 4.2(e) shows the distribution of hand movements on the lateral direction for adaptation trials measured at 100ms after the late perturbation time in each block/condition. In EARLY blocks, the distribution shows the response to the already perturbed target, in which case the shape is close to a Gaussian distribution with the mean approaching the final target position. This unimodal distribution is replaced by a tri-modal one in COG (left peak not shown), in which case the corrections were based on predictions alone and could not be adjusted on-line using the visual target. The higher peak on the right indicates that predictions in most trials were successful. In LATE, on the other hand, subjects were about to but have not reacted to the perturbations. Note the different distributions between DIRECT and SPRING. In the DIRECT condition, the single peak is close to 0, again indicates little or no learning; whereas the two peaks with equal probabilities in the SPRING condition reflect both

the use of prediction in on-line corrections and the uncertainty of such predictions.

Give the perturbation sequence was identical in all blocks/conditions, why learning effects were only observed in COG and LATE-SPRING? Was the lack of learning effect in other blocks/conditions because subjects didn't learn, or because subjects chose not to use what they learned? To answer this question, we tested subjects under the DIRECT condition using the EARLY block both before and after two learning blocks in Experiment 2. Each learning block was identical to the EARLY block in experiment 1, but had 10 cognitive trials interspersed among early perturbation trials to force subjects to learn to predict the perturbations. To measure the learning effect, the earliest correction time both before and after learning was estimated as before. After learning, the earliest correction time dropped from 120ms to 60ms, represented by the black and red dot respectively in Fig. 4.2(c). This change can not be simply due to the long exposure to the perturbations, since movements became very stable after 20 trials in EARLY in Experiment 1 (data not shown here). Therefore, on one hand, subjects used what they learned about the perturbations in on-line corrections once they were forced to learn, resulting in the drop of the earliest correction time. On the other hand, the fact that the earliest correction time after learning was still much later than that in COG indicates that predictions were not fully trusted when the target was visible to guide on-line corrections.

4.3.2 Predictions of optimal feedback control

The above results show that subjects were able to learn the statistical properties of external perturbations and apply such knowledge in on-line corrections to compensate for the delay of visual feedback. The learning effect, however, varied depending on the perturbation time and the hand cursor dynamics, regardless of the identical perturbation sequence. Is this behavior optimal? To answer this question, we model Experiment 1 using the Iterative Linear Quadratic Gaussian (ILQG) framework (see materials and methods for details).

To understand how corrective movements prior to perturbations influence the overall performance, we fix the control on the lateral direction before the perturbation time to a constant value, and compute the corresponding increase of cost, $\delta\mathbf{cost}$, comparing with the cost from the optimal solution where the control is free to change all the time. Fig. 4.3 reveals the change of $\delta\mathbf{cost}$ as a function of fixed initial control for different blocks/conditions. As can be seen, all the curves of $\delta\mathbf{cost}$ are smooth and convex around the corresponding minimum point. When

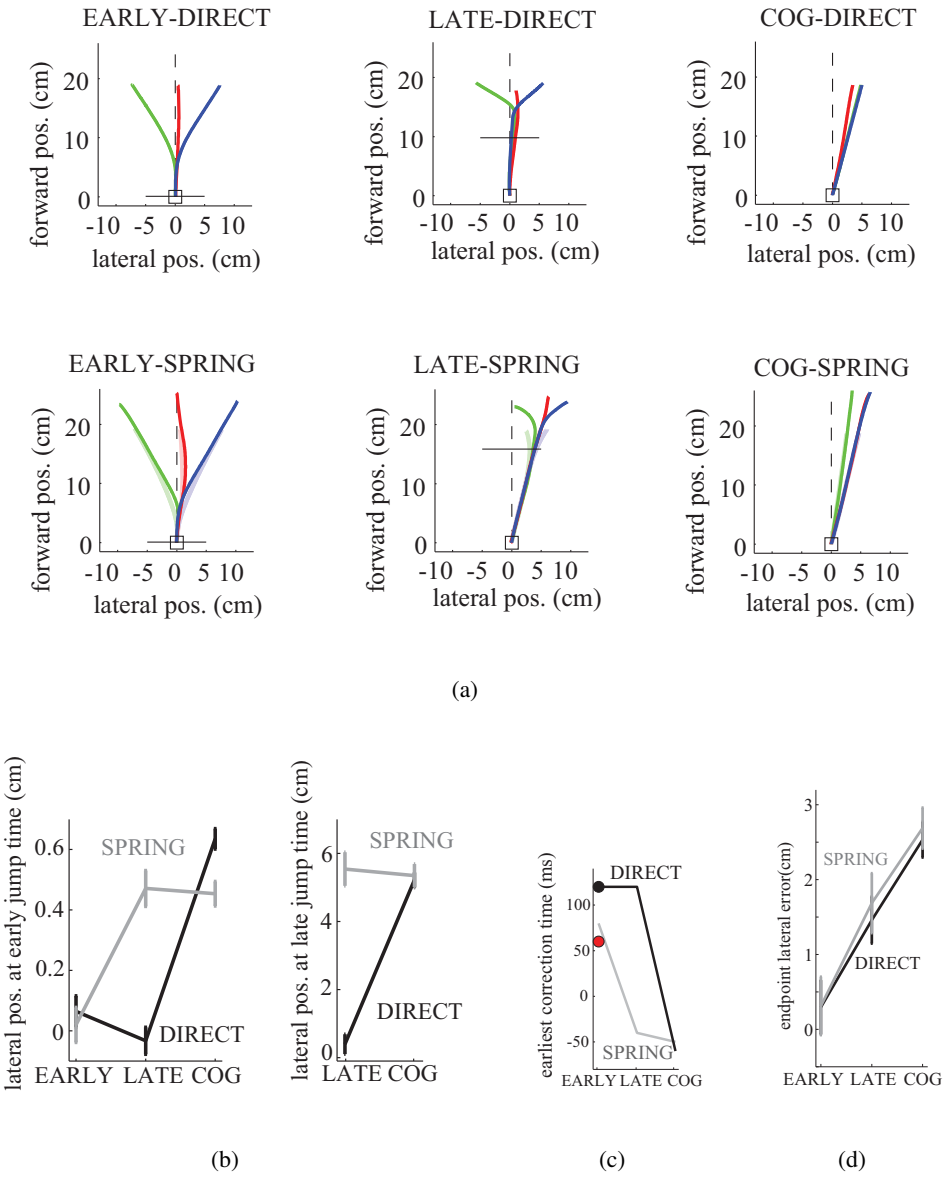
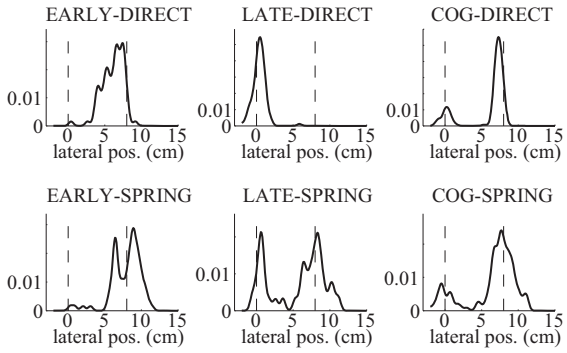


Figure 4.2 Learning results in response to target perturbations in Experiment 1 (to be continued).



(e)

Figure 4.2 Learning results in response to target perturbations in Experiment 1. **(a)**, Average hand paths for EARLY, LATE, COG in both DIRECT and SPRING conditions. Average has been done separately for baselines (catch no jump trials), trials where the target was perturbed to the same direction as in the previous trial (adaptation trials) or to the opposite direction (catch jump trials). Trials following left perturbations have been mirrored around the vertical axis and pooled with the corresponding trials following the right ones. Thus, the probability of going to the right is much larger. Blue: adaptation trials. Red: catch no jump trials. Green: catch jump trials. Dark, hand. Light, the hand cursor in SPRING. All movements started at the lower box. The dashed line connects the starting position and the initial target position. The horizontal line indicates the average hand position when perturbations occurred. **(b)**, Lateral hand position measured at 100ms after early and late perturbation time for only adaptation trials. **(c)**, Earliest correction time, measured as the first point in time where accelerations in left-perturbed trials are significantly different from those in right-perturbed trials, for adaptation trials only. To get an accurate measurement of accelerations, data from the accelerometer and the second order derivative of data from the Polhemus are combined optimally. Black dot: earliest correction time before learning in Experiment 2. Red dot, earliest correction time after learning in Experiment 2. **(d)**, Endpoint lateral errors for adaptation trials. **(e)**, Distribution of lateral hand positions measured at 100ms after late perturbation time for adaptation trials. Measurement from each trial is treated as a sample and samples from all trials are pool together to compute the histogram (unit interval is 0.1cm). Then the histogram is smoothed using 1d Gaussian kernels centered at each interval with variance of 0.1cm and nomalized so that its sum is 1. Dashed lines represent the initial and final target positions.

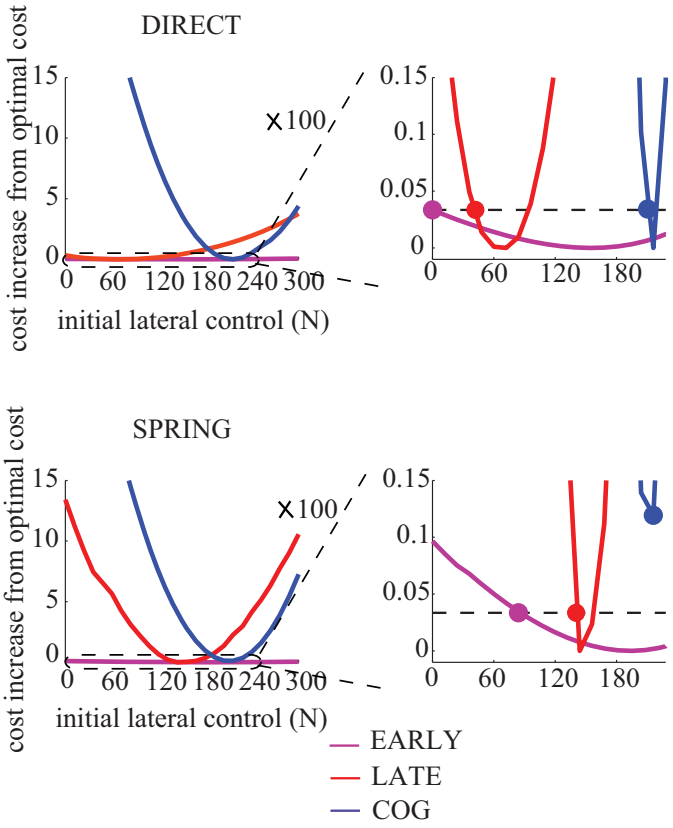
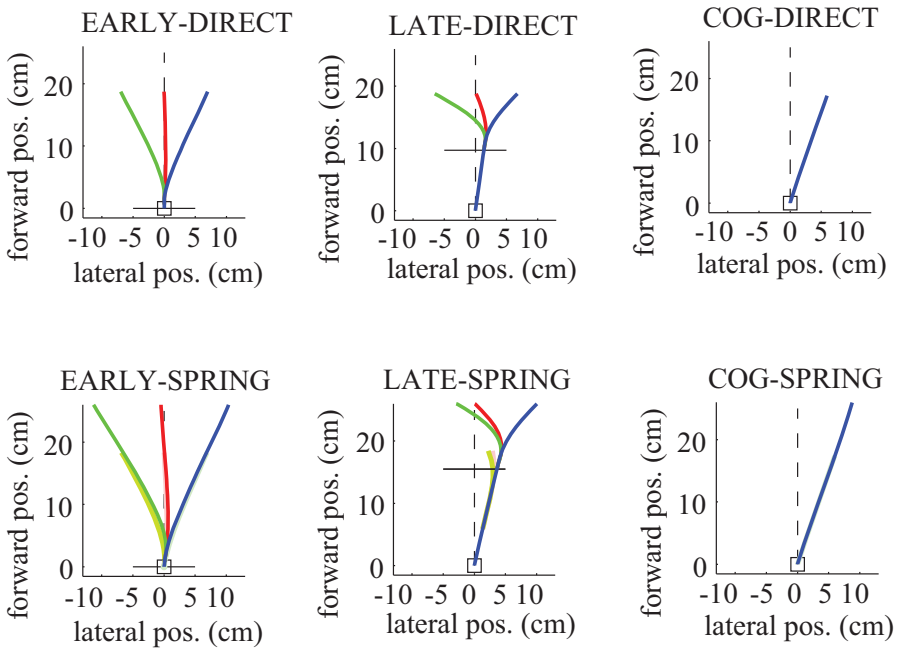


Figure 4.3 Increase of cost, compared with the cost from the optimal solution, as a function of fixed initial lateral control from the optimal feedback control model for Experiment 1. Right panels show the zoom-in results of the left. Filled circle represents the initial lateral control chosen to reproduce the experimental data for each block/condition.

the lateral control is fixed to 0, or equivalently no correction is made before the perturbation, $\delta\mathbf{cost}$ in COG and LATE-SPRING is much larger than the rest ($\delta\mathbf{cost}$ >10), whereas $\delta\mathbf{cost}$ in EARLY, especially in the DIRECT condition, becomes negligible ($\delta\mathbf{cost}$ <0.1). That is, performance is reduced without using prediction-based on-line corrections, but the impairment is much larger in COG and LATE-SPRING than the rest. This explains why learning effects were significant only in COG and LATE-SPRING. The sensitivity of performance to corrections prior to perturbations is also revealed by the slope of $\delta\mathbf{cost}$. The much more flat curve in EARLY implies that subjects' reluctance to use predictions would have little impact on their performance, which explains why little or no learning effect was observed in EARLY in the experiment. If subjects would like to sacrifice a little performance to rely more on the visual feedback rather than the prediction in on-line corrections in EARLY, they should do the same thing in other cases. Based on this assumption, we draw a line through $\delta\mathbf{cost} = 0.03$, which goes across each curve twice, except for COG-SPRING, one before the minimum point and one after. Fig. 4.4(a)-(d) shows the results when the lateral control before perturbations is fixed to the corresponding left cross for each block/condition separately (lateral control in COG-SPRING is fixed to the point corresponding to the lowest $\delta\mathbf{cost}$), in the same format as in Fig. 4.2(a)-(d). Note the close correspondence with experimental data, especially the much larger and earlier corrections in COG and LATE-SPRING. If the model is free to choose lateral control before perturbations to achieve the lowest cost, the resulting corrective movements are close to data except for the bigger predicted corrections in EARLY (see Fig. 4.4(e). Therefore, as long as the performance was good enough, subjects preferred suboptimal strategies to avoid using prediction in on-line corrections in Experiment 1.

4.3.3 On-line corrections to perturbations generated from different distributions

Why did subjects prefer suboptimal control strategies instead of fully using prediction in on-line corrections? One possible reason is that since target was displaced both right and left, moving to one direction before the perturbation could increase the distance to the target if the perturbation was to the opposite direction. In this case, the large error had to be corrected using large effort, which subjects may try to avoid. If the motor system is indeed sensitive to large energy consumptions, limiting the perturbation to only one direction should increase the amount of prediction-based corrections. This hypothesis was tested in Experiment 3, where



(a)

Figure 4.4 An optimal feedback control model for Experiment 1 (to be continued).

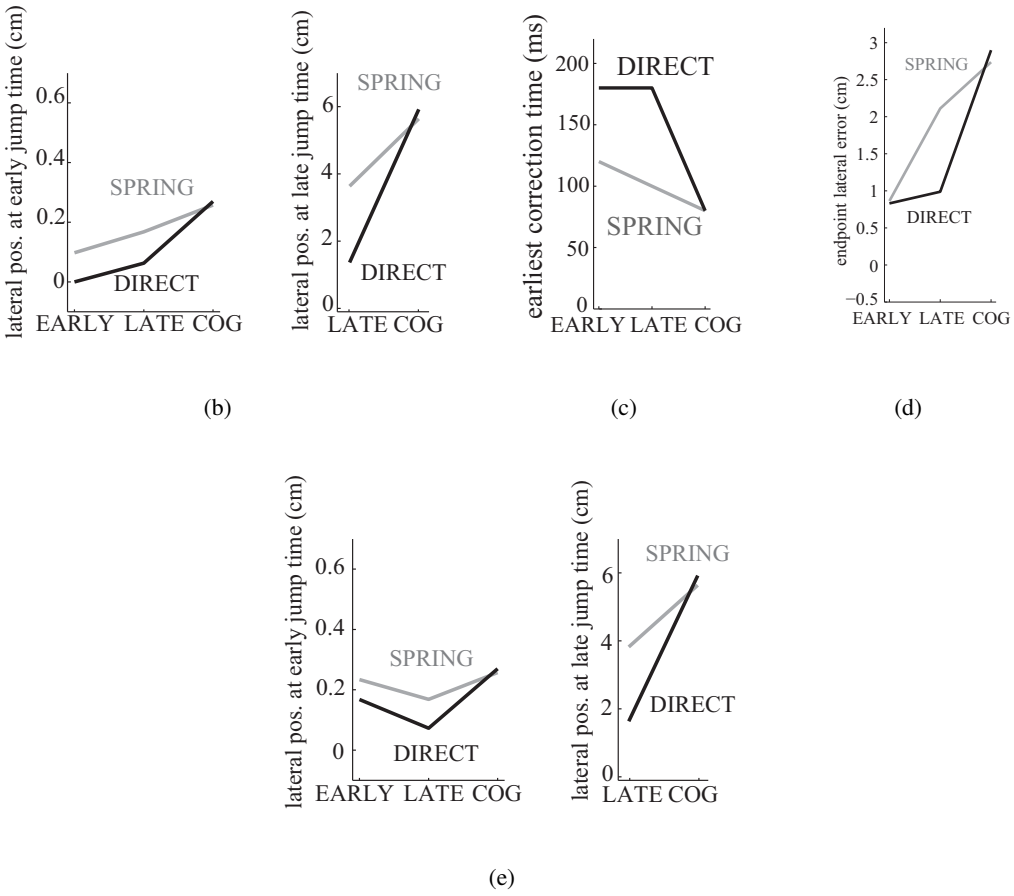


Figure 4.4 An optimal feedback control model for Experiment 1. **(a)-(d)**, Same as the corresponding plots of Fig. 4.2(a)-(d), but for data generated by the suboptimal solution from the optimal feedback control model. Earliest correction time from the model is measured as the first point in time where the lateral velocity exceeds 3.5cm/m. Fig. 4.2(e)same as 4.2(b) but based on the optimal solution from the model.

perturbations were independently generated from either a bimodal distribution (bimodal block) or a tri-modal distribution (tri-modal block). The two distributions shared the same mean but had different shapes, especially, perturbations could occur both left and right in the tri-modal one but only to the right in the bimodal one.

We first compare the hand paths in the two blocks within each subject, and find that the difference varied slightly depending on which block was experienced first. To get rid of the interference from the past, only data in the first experienced block for each subject are used. Fig. 4.5(a) left panel shows the average hand paths in the two blocks, filled area indicates where corrections to right perturbations in the two blocks are significantly different. As we expected, initial hand movements were away from the middle and moved towards the right, which had a higher probability for perturbations to occur. Furthermore, corrections prior to perturbations became significantly larger once the perturbation was limited to the right as in the bimodal block, also shown in Fig. 4.5(b) left panel, confirming the sensitivity of the motor system to large energy consumptions.

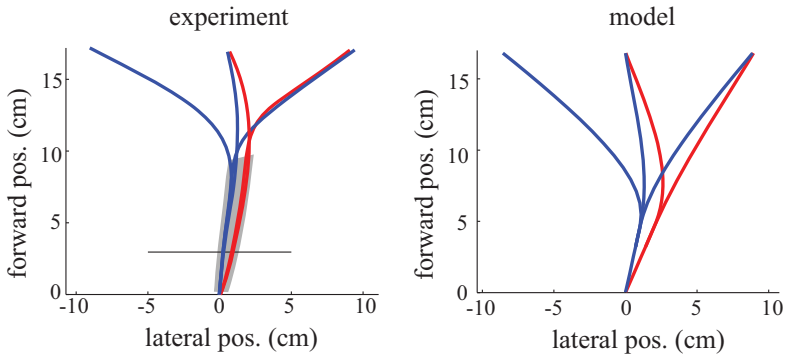
We also model Experiment 3 in the same ILQG framework as before, and the results are plotted in the same format as data on the right in Fig. 4.5(a)-(b). When the energy cost is framed as a quadratic term of control, widely used due to its nice mathematical property, corrections prior to perturbations in the two blocks are identical. The discrepancy occurs only when the order of control goes higher, shown in the right panel in Fig. 4.5(a). Results in the right panel in Fig. 4.5(a) are based on the 4^{th} order energy cost. As can be seen, model prediction faithfully captures the different corrective movements prior to perturbations between the two distributions.

4.3.4 Changes in on-line corrections as a result of learning

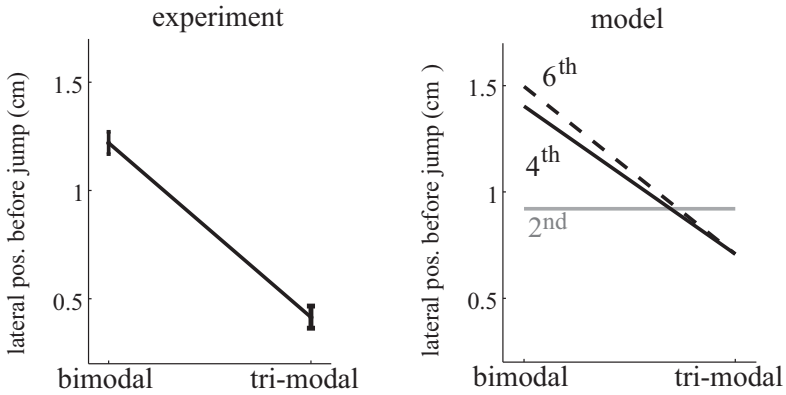
The above analysis focuses on on-line corrections to perturbed target after learning. In this section, we further explore how on-line corrections changed during learning.

We first quantify learning in experiment 1 using a version of the statespace approach following the idea of trial-to-trial learning as follows [71]:

$$z(n+1) = az(n) + B\mathbf{w}(n) + \eta(n)$$
$$y(n) = z(n) + \gamma(n)$$



(b)



(c)

Figure 4.5 Learning results in response to perturbations in Experiment 3 (left column), and the corresponding prediction by the optimal feedback control model (right column). **(a)**, Hand paths in bimodal and tri-modal blocks in Experiment 3. Red, bimodal. Blue, tri-modal. In the left panel, filled area indicates that the paths in right-perturbed trials are significantly different between the two blocks. In the right panel, energy cost was modeled as a forth order term of the control to get the result. **(b)**, Lateral hand position measured at 100ms after the perturbation time in the two blocks. In the right panel, energy cost is modeled as a second, forth, and sixth order term of the control respectively.

The correction on trial n is denoted by $y(n)$, representing the position in the corrective direction from the average unperturbed trajectory measured at 100ms after the perturbation. Thus, the deviation from the average unperturbed trajectory should be due to prediction. $\mathbf{w}(n)$ has two elements specifying the endpoint target position and the endpoint positional error (again in the corrective direction). $z(n)$ is the internal learning state that the model is free to use in whatever way is needed to fit the data. $\eta\,(n)$ and $\gamma\,(n)$ are independent zero-mean one-dimensional Gaussian white noise processes with variances q and r. Given these measurements, the most likely values of a, B, q, r as well as the sequence of learning states $z\,(n)$ are computed using the expectation maximization (EM) algorithm [76, 71]. Because EM can get trapped in local minima, model fitting is run multiple times with different initial conditions and the best result is used. We fit the model separately for the EARLY, LATE and COG block as well as the first RANDOM block for each subject. To evaluate the performance of the learning model, we regress $y(n)$ on $z\,(n)$. To measure the contribution of learning from previous perturbation and previous error alone, we fit the model again using either one of the two and regressed $y(n)$ on the fitted $z\,(n)$ separately. The R^2 and p values for the regressions are averaged over subjects for each block under DIRECT and SPRING conditions separately. Table 2 shows the average R^2 values multiplied by 100 (to obtain a measure of variance explained), only for the cases in which $p < 0.01$ on average. For the RANDOM block in both conditions, the regression is not significant even on the full model, indicating no learning effect, consistent with previous report [77, 71]. When the perturbation was not random as in EARLY, LATE and COG, the regressions, either on the full model or previous perturbation alone or previous error alone, are significant in all blocks/conditions except for EARLY in the DIRECT condition, similar to the result from section 1. Furthermore, R^2 for using previous perturbation alone is in general bigger than using previous error alone. In fact, the combined model is only slightly but systematically better than using previous perturbation alone. This suggests that instead of naively correcting the previous error as observed in random hand perturbation experiments, subjects learned the pattern of perturbations applied to the target, and used such knowledge in subsequent movements.

However, trial-to-trial learning would not be effective in Experiment 1, where perturbations followed a complex repeat-and-switch pattern. Once the target was perturbed to a certain direction, the same perturbation would repeat in the next 2-4 trials in a row and then switch to the other direction and repeat. To understand whether subjects learned this repeat-and-switch

Table 4.3 $100R^2$ for regressions on fitted model

	DIRECT				SPRING			
regression	RANDOM	EARLY	LATE	COG	RANDOM	EARLY	LATE	COG
full	·	·	45*	50*	·	23*	39*	42*
previous pert.	·	·	42*	50*	·	20*	35*	40*
previous error	·	·	34*	50*	·	13*	34*	32*

pattern, we analyzed sequence of trials where all the perturbations were to the same direction in a row. Trials from different sequences were grouped according to the number of preceding trials with the same perturbations. For example, 0 represents a different perturbation from the previous trial, corresponding to catch jump trials; 1-4 represent trials where the same perturbation has repeated for 1-4 times, corresponding to adaptation trials. Catch no jump trials are not analyzed here. Fig. 4.6(a) compares the initial correction, defined as the lateral hand position before reacting to perturbations (trials in COG are measured at the same time as in LATE), averaged in the first 20 trials (grey) and last 20 trials (black) for EARLY, LATE and COG separately in the SPRING condition. In EARLY, initial corrections were never significantly different from 0, confirming the lack of learning again. In LATE and COG, in contrast, initial corrections increased significantly after learning. Although the corrections after catch jump trials and after adaptation trials were similar before learning, after learning, corrections after adaptation trials became much larger than those after catch jump trials. This dramatic increase after catch jump trials implies that subjects followed the most recent perturbation while ignoring the perturbations experienced in the past. This resetting mechanism can not be accounted for using a fixed learning rate, as assumed in trial-to-trial learning. In fact, we fit the trial-to-trial learning model described above to the first and last 20 trials in each corresponding block separately, and analyze the best fitted initial correction $z\,(n)$ the same way as we analyze the data, shown in blue lines in Fig. 4.6(a). As we can see, the trial-to-trial model can not reproduce the sharp increase of the initial corrections after catch jump trials. Thus, subjects learned the perturbation pattern in a more complex way than the trail-to-trial manner.

Fig. 4.6(a) also reflects a different learning mechanism between the LATE and COG blocks in the SPRING condition, despite the similar learning effects between the two from section 1. In LATE, initial corrections to adaptation trials were similar, regardless of the times the

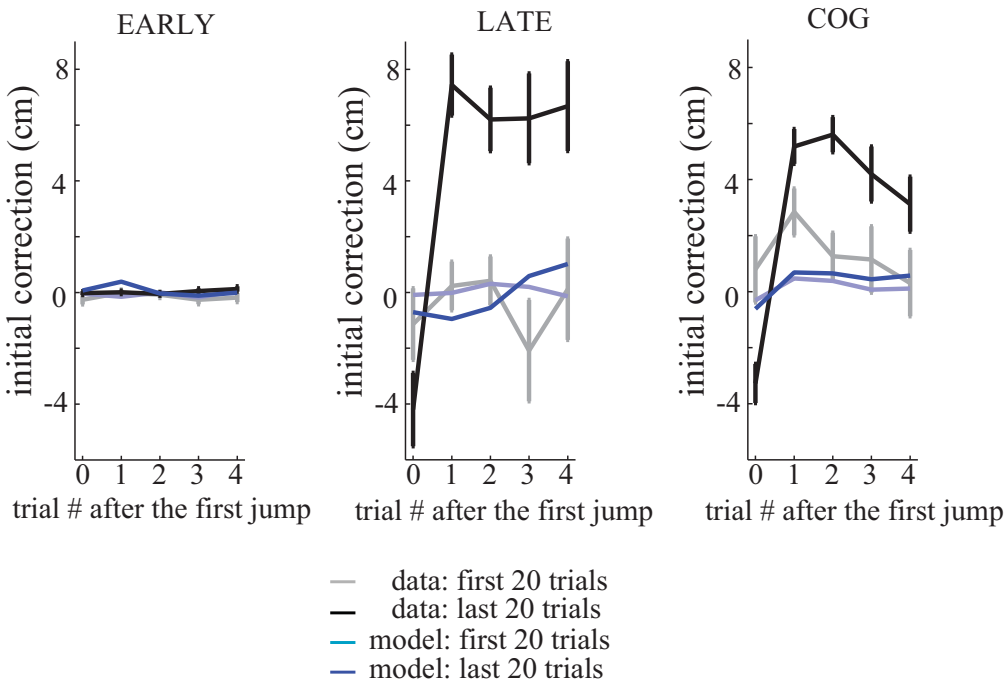
same perturbation has repeated. In COG, in contrast, initial corrections dropped significantly after 4 trials in a row with the same perturbation. As subjects reported after the experiment, they noticed not only the repeat-and-switch pattern, but also the constraints on the number of repetitions in this case. This counting strategy implies that learning became even more complex once target was not visible to guide on-line corrections.

To investigate what caused the different learning effects under the same perturbation sequence in Experiment 1, we examined the relationship between the endpoint lateral error and the corresponding initial correction through adaptation trials. Here we only focus on the LATE block because of the salient difference in the corrective movements between DIRECT and SPRING from section 1, and the results are plotted in Fig. 4.6(b). Despite the same perturbation sequence and the same perturbation time in the two conditions, significant increase of the initial correction through learning are only observed in SPRING, causing the endpoint lateral error to drop significantly. As a result, although the task was much harder when the hand cursor was delayed in SPRING, as reflected from the much larger error in the beginning of the block in this case, the error was not significantly different from that in the DIRECT condition in the end of the block. This result suggests that prediction was used in on-line corrections only if it could lead to accuracy improvement. Therefore, the reason why no learning effect was observed in EARLY might be due to the fact that full corrections could be achieved by relying on feedback alone.

The relationship between the endpoint lateral error and the corresponding initial correction is also analyzed to data in the bimodal block in Experiment 3, shown in Fig. 4.6(c). Different from Experiment 1, the significant increase of initial correction didn't reduce the endpoint lateral error. This implies that in addition to the endpoint accuracy, the motor system is also sensitive to other factors. One possibility is the energetic efficiency, as suggested by results in section 3.

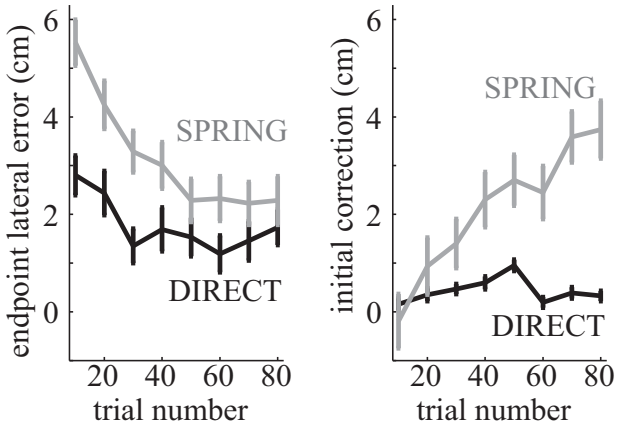
4.4 Discussion

This paper studies how subjects learned the statistical properties of target perturbations and applied such knowledge in on-line corrections. Our central argument is that people learn from the changing world in a flexible way. This is supported by two lines of evidence. First,

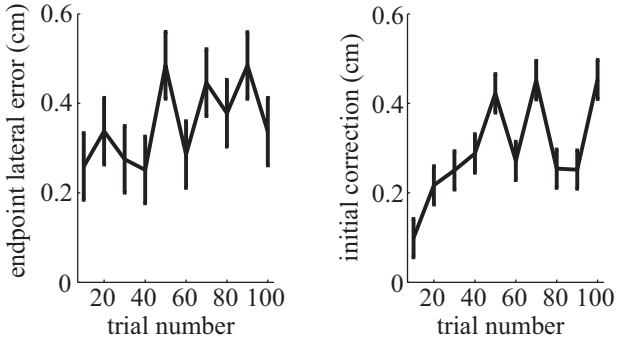


(a)

Figure 4.6 Learning process during Experiment 1 and 3 (to be continued).



(b)



(c)

Figure 4.6 Learning process during Experiment 1 and 3. **(a)**, Initial correction, defined as the lateral hand position at 100ms after perturbations (trials in COG are measured at the same time as in LATE), in the first 20 trials (light) and the last 20 trials (dark) in the SPRING condition from both experimental data (black) and the fitted trial-by-trial learning model (blue). Results are averaged according to how many times the same perturbation has repeated in the preceding trials in a row. More specifically, 0 represents the catch jump trials, 1-4 represent the adaptation trials, and catch no jump trials are not shown. **(b)**, Endpoint lateral error and the corresponding initial correction during adaptation trials in LATE in both DIRECT and SPRING conditions in Experiment 1. Results are averaged every 10 trials for all subjects in the group. **(c)**, Endpoint lateral error and the corresponding initial correction during learning in the bimodal block in Experiment 3. Results are averaged every 10 trials for all subjects in the group.

the same perturbation sequence could elicit different learning effects, depending on whether the use of prediction in on-line corrections could lead to better endpoint positional accuracy and less energy consumptions. Second, instead of adopting a fixed learning rule, subjects used different learning mechanisms when the perturbations followed a more complex pattern.

4.4.1 Different learning effects in on-line corrections under the same perturbation sequence

One main phenomena we analyzed, different learning effects in on-line corrections in response to the same perturbation sequence under different visual feedback conditions in Experiment 1, shows that although the motor system is able to learn the statistics of a changing environment from the experience, learning effects are more significant when the sensory feedback is not sufficient to achieve good performance. The correlation between the initial correction and the corresponding endpoint lateral error indicates that the motor system is sensitive to endpoint positional accuracy, and will use prediction in on-line corrections only when it could improve such accuracy. The fact that limiting the perturbation to only one direction instead of two reduced subjects' reluctance to use prediction in on-line corrections in Experiment 3 suggests that the motor system is also sensitive to energetic efficiency, and will not rely on a prediction if it has the potential to lead to large energy consumptions. This may be the first demonstration that the output of the visuomotor feedback loop is not a simple readout of the prediction. Instead, how adjusting motor commands based on a prediction will affect the overall performance is also taken into account, with emphasis given to the uncertainty of the prediction.

Such flexible adaptive response based on the statistical properties of a changing environment is inherent in our optimal feedback control model. It represents performance using a cost function and at each time step computes the control law to minimize the expected final cost by taking into account both uncertainties from the changing world and the noise and delay in the sensorimotor system. In this way, our model is able to compare how different control strategies based on the same prediction affect the overall performance, always choosing the optimal strategy. To explain the data in the experiments, we used a composite cost function encouraging both endpoint positional accuracy and energetic efficiency. Moreover, to model the sensitivity to large energy consumptions, the traditional quadratic energy cost was replaced by a higher order term. Although this non-quadratic cost as well as the non-Gaussian noise in modeling perturbations

can not be dealt with under the LQG framework, they were successfully accommodated by the ILQG method.

4.4.2 High level learning mechanisms in response to non-random perturbations

It has been suggested that motor learning occurs in a trial-to-trial manner, which, however, is apparently not effective when the environment is changing in complex ways, such as in the real world. In Experiment 1, we demonstrated that different levels of learning mechanisms may be involved in learning the repeat-and-switch pattern of perturbations. Some of them are more complex than trial-to-trial learning. When the visual feedback was sufficient to make full corrections as in EARLY-DIRECT, subjects didn't learn the perturbation pattern at all. When learning was necessary to compensate for delays of the perturbation and the hand cursor as in LATE-SPRING, subjects moved towards the most recent perturbation direction in a predictive way regardless of the past, and the learning rate in catch jump trials was much smaller than that in adaptation trials. This result can not be account for by trial-to-trial learning, which assumes accumulating learning with a constant learning rate. Finally, when learning was the only means to make corrections as in COG, subjects moved to the opposite direction to the previous perturbation after seeing the same perturbation for several times in a row. As they reported after the experiment, they noticed not only the repeat-and-switch pattern, but also the constraints on the number of repetitions. This counting strategy further challenges trial-to-trial learning, which relies on the idea that the neural structures modified as a result of motor learning don't explicitly retrain memories of perturbations beyond one trial in the past.

4.4.3 Explanations to the lack of learning effect in response to random perturbations

Based on the above results, we propose three explanations for the apparent lack of trial-by-trial learning in some target perturbation experiments. First, subjects always learn to predict but do not use the prediction unless the lack of time for on-line corrections forces them to do so. In this way they can still accomplish the task, while avoiding large corrections in cases where the perturbation happens to be opposite to the expected direction. Second, subjects do not bother to learn in conditions where they know they can rely on on-line corrections. Third, the

motor system has a preference for moving towards the current target, and needs a good reason (such as task failure) to deviate from this preferred strategy.

Acknowledgment

This chapter, in part, has been submitted for publication in *the Journal of Neuroscience*. The dissertation author was the primary author of the paper.

Chapter 5

Hierarchical Control as an Approximation to Optimal Control on a Realistic Arm Model

5.1 Introduction

The human body has over 250 muscles each with a distinct mechanical action at one or more joints, and the nervous system must learn which muscles to use to perform a movement [62]. key to understanding biological motor control is thus to solve the control problem of complex redundant systems. This control problem, however, has not been well addressed because the the nonlinear dynamics and high-dimensional state and control spaces of human body prevent the use of many traditional methods for controller design.

As an attempt to control a redundant manipulator, [7] proposed a hierarchical control framework. This framework is inspired by two observations. First, from a computational viewpoint, optimal feedback controllers for redundant systems exhibit hierarchical organizations [4, 16]. Secondly, from a biological viewpoint, it is known that sensorimotor control occurs simultaneously on many levels [39, 40]. Lower-level circuits (e.g., the spinal cord) interact with the musculoskeletal system directly by both receiving rich sensory input and generating corresponding motor outputs before the rest of the brain has had time to react to that input. High-level circuits (e.g., the motor cortex), on the other hand, operate on a more abstract and goal-related

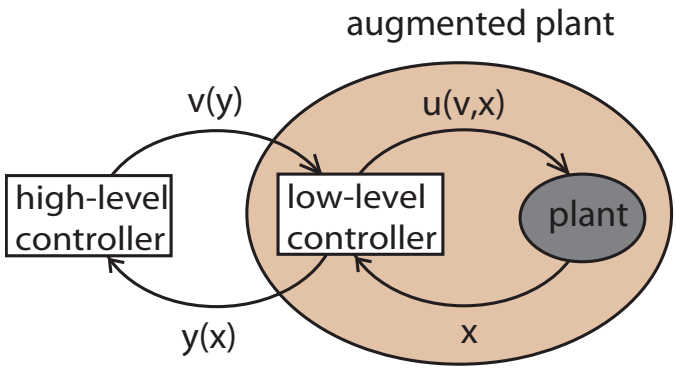


Figure 5.1 Schematic illustration of the hierarchical control framework [7].

movement representations [41]. The proposed hierarchical framework also has two layers, illustrated in Fig. 5.1. The plant is augmented with a low-level feedback controller, which receives information about the plant state $\mathbf{x}$, and sends to the high-level $\mathbf{y}(\mathbf{x})$ that captures the task-relevant aspects of plant dynamics but has reduced dimensionality. The high-level monitors task progress, and issues commands $\mathbf{v}(\mathbf{y})$ to specify how $\mathbf{y}(\mathbf{x})$ should change to archive the goal. Then the low-level controller computes energy-efficient controls $\mathbf{u}(\mathbf{v}, \mathbf{x})$ to control the plant to accomplish the trajectory designed from the high-level. In this way, the high-level solves the optimal control problem without considering all the details of the plant and thus avoid running into the curse of dimensionality, the low-level performs an instantaneous feedback transformation to deal with the details. This hierarchical control framework has been tested on a simplified arm model with 2 degrees of freedom (DOF) in 2D movements. The goal of this paper is to apply this framework to a more realistic arm model with 7 DOF and 14 muscles on 3D movements. Our focus is on designing the high-level controller to capture the low-level dynamics, and designing the low-lever controller to generate biologically plausible arm configurations to accomplish the control designed by the high-level.

5.2 General framework

Mathematically, the hierarchical control framework can be described as

$$\text{low-level dynamics} \quad \dot{\mathbf{x}}(t) = \mathbf{a}(\mathbf{x}(t)) + B(\mathbf{x}(t))\mathbf{u}(t) \tag{5.1}$$

$$\text{high-level dynamics} \quad \dot{\mathbf{y}}(t) = \mathbf{f}(\mathbf{y}(t)) + G(\mathbf{y}(t))\mathbf{v}(t) \tag{5.2}$$

$$\text{low-level to high-level} \quad \mathbf{y} = \mathbf{h}(\mathbf{x}) \tag{5.3}$$

where

	low-level	**high-level**
state vector	$\mathbf{x} \in \mathbb{R}^{n_x}$	$\mathbf{y} \in \mathbb{R}^{n_y}$
control vector	$\mathbf{u} \in \mathbb{R}^{n_u}$	$\mathbf{v} \in \mathbb{R}^{n_v}$
passive dynamics	$\mathbf{a}(\mathbf{x})$	$\mathbf{f}(\mathbf{y})$
control-dependent dynamics	$B(\mathbf{x})\mathbf{u}$	$G(\mathbf{y})\mathbf{v}$

Here function $\mathbf{h}$ defines the transformation from low-level to high-level circuits. Such transformation should allow the high-level to acquire enough information in computing the state-dependent cost $q(t, \mathbf{x})$, and at the same time operate on a reduced state space to avoid running into the curse of dimensionality during optimization. In other words, $\mathbf{h}$ needs to satisfy $\exists \tilde{q}$ s.t. $\tilde{q}(t, \mathbf{y}) = q(t, \mathbf{x})$ and $n_{\mathbf{y}} < n_{\mathbf{x}}$. The goal of the high-level controller is to choose $\mathbf{v}$ to accomplish the goal of the task without considering the full details of the plant.

Differentiating Eq. 5.3 w.r.t. time t and using Eq. 5.1, the dynamics of $\mathbf{y}$ become

$$\dot{\mathbf{y}} = H(\mathbf{x})(\mathbf{a}(\mathbf{x}) + B(\mathbf{x})\mathbf{u}) \tag{5.4}$$

where $H(\mathbf{x}) = \partial \mathbf{h}(\mathbf{x}) / \partial \mathbf{x}$ is the Jacobian of the function $\mathbf{h}$. The goal of the low-lever controller is to choose $\mathbf{u}(\mathbf{v}, \mathbf{x})$ so that the $\mathbf{y}$ dynamics from the low-level (Eq. 5.4) match those from the high-level (Eq. 5.2). That is

$$H(\mathbf{x})\mathbf{a}(\mathbf{x}) + H(\mathbf{x})B(\mathbf{x})\mathbf{u} = \mathbf{f}(\mathbf{y}) + G(\mathbf{y})\mathbf{v} \tag{5.5}$$

The goals of the two levels can be archived using two methods:

1. Explicit modeling

   We can treat the high-level as an autonomous system, explicitly model its dynamics, and

then use standard optimization techniques to solve the optimal control problem in the high-level. The low-level then seeks energy efficient $\mathbf{u}$ to satisfy Eq. 5.5. Ideally, we would like to have the high-level dynamics to mimic those from the low-level, so

$$\mathbf{f}(\mathbf{y}) \approx H(\mathbf{x})\mathbf{a}(\mathbf{x}) \tag{5.6}$$

However, when the high-level considers only the end-effector location as in most reaching movements, Eq. 5.6 cannot be guaranteed because different arm configurations may result in the same end-effector location. In other words,

$$H(\mathbf{x_1})\mathbf{a}(\mathbf{x_1}) \neq H(\mathbf{x_2})\mathbf{a}(\mathbf{x_2}) \text{ even } \mathbf{h}(\mathbf{x_1}) = \mathbf{h}(\mathbf{x_2}) \tag{5.7}$$

In this case, low-level dynamics cannot be fully captured by the high-level. As a result, the optimal solution generated from the high-level may not be optimal with respect to the low-level dynamics.

2. Implicit modeling

   The drawback of explicit modeling can be avoided by not modeling the passive dynamics explicitly on the high-level. Instead, the high-level controller is given on-line access to $H(\mathbf{x})\mathbf{a}(\mathbf{x})$ and is responsible for dealing with it. The discrepancy between $\mathbf{f}(\mathbf{y})$ and $H(\mathbf{x})\mathbf{a}(\mathbf{x})$ is no longer compensated on the low-level. So the low-level controller only needs to satisfy

   $$H(\mathbf{x})B(\mathbf{x})\mathbf{u} = G(\mathbf{y})\mathbf{v} \tag{5.8}$$

   In this way, the high-level dynamics can exploit the passive dynamics of the plant while operating on a lower-dimensional system. In addition, now we can apply Eq. 5.8 to match the low-level control cost $r(\mathbf{u}, \mathbf{x})$ exactly to the high-level $\tilde{r}(\mathbf{v}, \mathbf{y})$. In this way, the high-level can seek to solve the exact optimal control problem with respect to the true plant dynamics. However, since now the high-level is no longer an autonomous system, regular optimization tools are not guaranteed to provide optimal solutions.

5.3 Low-level dynamics

Here we considering a 7-DOF arm with the same skeletal structure as human arms: the shoulder is modeled as a 3-DOF joint (only the glenohumeral joint was taken into account

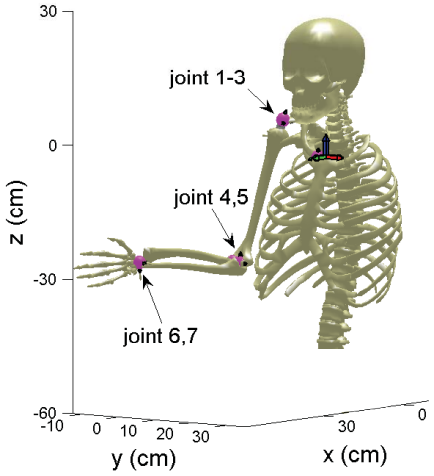


Figure 5.2 A realistic arm model.

here), with flexion-extension, abduction-adduction and external-internal rotation; the elbow is modeled as a 2-DOF joint (humeroulnar joint and radioulnar joint), with flexion-extension and pronation-supination movements; the wrist is modeled as a 2-DOF joint, with flexion-extension and abduction-adduction movements.

The low-level dynamics are modeled as a third-order system resembling those in human arms [108]. The state vector of the low-level includes joint angles (θ), joint velocities ($\dot{\theta}$) as well as muscle activations ($\mathbf{s}$). Although each joint is actually controlled by a groups of muscles which are also connected to other joints, here we consider a simplified case where each joint is independently controlled by two muscles acting on opposite directions. So the complete system has a 28D state vector $\mathbf{x} \doteq (\theta_1, ..., \theta_7, \dot{\theta}_1, ..., \dot{\theta}_7, s_1, ...s_{14})^T$ and 14D control vector $\mathbf{u} \doteq (u_1, ..., u_{14})^T$.

Forward dynamics of the arm can be expressed as

$$\ddot{\theta} = I(\theta)^{-1} \left(\tau(\theta, \dot{\theta}, \mathbf{s}) - \mathbf{n}(\theta, \dot{\theta}) \right) \tag{5.9}$$

where $I(\theta) \in R^{7\times7}$ is a positive definite symmetric inertia matrix, $\tau(\theta, \dot{\theta}, \mathbf{s}) \in R^7$ represents joint torques, and $\mathbf{n}(\theta, \dot{\theta}) \in R^7$ are Centripetal, Coriolis, gravitational, and viscoelastic forces.

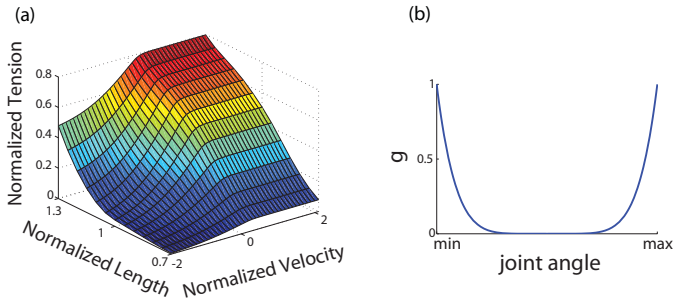


Figure 5.3 **(a)**, The total force produced by a muscle fiber at a given level of activation depends on both its instantaneous length and velocity, which are independent kinematic variables. **(b)**, polynomial function to punish moving towards joint limits.

Joint torques are generated by muscles following

$$\tau(\theta, \dot{\theta}, \mathbf{s}) = M(\theta)T(\mathbf{s}, \mathbf{l}(\theta), \dot{\mathbf{l}}(\theta, \dot{\theta})) \tag{5.10}$$

where $M(\theta) \in \mathbb{R}^{7\times 7}$ is the moment arm, defined as the perpendicular distance from the muscle's line of action to the joint's center of rotation. Although moment arms are posture dependent, here we consider it as constant. $T(\mathbf{s}, \mathbf{l}(\theta), \dot{\mathbf{l}}(\theta, \dot{\theta}))$ represents muscle tension. The tension produced by muscle i depends on its physiological cross-sectional area (PCSA) and activation state s_i, as well as the muscle length $\mathbf{l}$ and velocity $\dot{\mathbf{l}}$. The substantial length-and-velocity dependence is modeled based on the Virtual Muscle model [109] which provides a state-of-the-art fit to a range of physiological data, with slight simplification for computational purpose, illustrated in Fig. 5.3**(a)** for maximal activation $s_i = 1$. Muscle activation states $\mathbf{s}$ have first-order low-pass filter dynamics

$$\dot{s_i} = \frac{1}{\alpha}(u_i - s_i) \tag{5.11}$$

with $\alpha = 40msec$. Each control u_i is constrained to the range $[0, 1]$. In this way, our arm model captures most key features of human arms even with some simplifications.

The low-level dynamics described above can be summarized as follows:

$$\mathbf{a}(\mathbf{x}) = \begin{bmatrix} \dot{\theta} \\ I(\theta)^{-1}\left(\tau(\theta, \dot{\theta}, \mathbf{s}) - \mathbf{n}(\theta, \dot{\theta})\right) \\ -\frac{1}{\alpha}\mathbf{s} \end{bmatrix}, B(\mathbf{x}) = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \frac{\mathbf{1}}{\alpha} \end{bmatrix} \tag{5.12}$$

5.4 Design of high-level controller

The goal of movement in everyday life (e.g. pick up a cup of coffee) is usually defined in Cartesian hand coordinates rather than in joint coordinates, and evidence has also shown that human movements are planned in Cartesian coordinates [3]. Therefore, we only include 3D positions ($\mathbf{p}$) and velocities ($\dot{\mathbf{p}}$) of the end-effector (fingertip) in Cartesian hand coordinate in the high-level, so $\mathbf{y} \doteq [p_x, p_y, p_z, \dot{p}_x, \dot{p}_y, \dot{p}_z]^T$. The high-level control, accordingly, is defined as 3D forces $\mathbf{v} \doteq [v_x, v_y, v_z]^T$. Including velocity in the high-level is necessary, because it has been shown that people would slow down during a simple goal-directed reaching movement even the only requirement was the endpoint accuracy [71], and [7] also demonstrated that controlling hand position through instantaneous velocity commands are too far removed from muscles that have to carry them out.

As introduced in Section 5.2, there are two ways to design the high-level controller. Using the explicit modeling, we can simplify the entire arm into a point mass (m) and model the dynamics as a simple linear system:

$$\mathbf{f}(\mathbf{y}) = \begin{bmatrix} \dot{\mathbf{p}} \\ \mathbf{0} \end{bmatrix}, G(\mathbf{y}) = \begin{bmatrix} \mathbf{0} \\ \frac{1}{m} \end{bmatrix} \tag{5.13}$$

The transformation from low-level to high-level can be represented as

$$H(\mathbf{x}) = \begin{bmatrix} J(\theta) & \mathbf{0} & \mathbf{0} \\ \dot{J}(\theta) & J(\theta) & \mathbf{0} \end{bmatrix} \tag{5.14}$$

where J is the Jacobian $J(\theta) = \partial \mathbf{p} / \partial \theta$, and $\dot{J}(\theta) = \frac{d}{dt}(J(\theta))$.

We can also use the implicit modeling. Instead of explicitly modeling the high-level dynamics, we add a constant vector in the high-level states $\mathbf{y} \doteq [p_x, p_y, p_z, \dot{p}_x, \dot{p}_y, \dot{p}_z, 1, 1, 1]^T$ to capture the low-level dynamics:

$$\mathbf{f}(\mathbf{y}) = \begin{bmatrix} \dot{\mathbf{p}} \\ \mathbf{e} \\ \mathbf{0} \end{bmatrix}, G(\mathbf{y}) = \begin{bmatrix} \mathbf{0} \\ I(\mathbf{y})^{-1} \\ \mathbf{0} \end{bmatrix} \tag{5.15}$$

where $\mathbf{e}$ is the 4^{th} to 6^{th} rows of $H(\mathbf{x})\mathbf{a}(\mathbf{x})$. $I(\mathbf{y})$ is the end-effector inertial matrix and

$I(\mathbf{y})^{-1} = J(\theta)I(\theta)^{-1}J(\theta)^T$. The transformation from low-level to high-level becomes

$$H(\mathbf{x}) = \begin{bmatrix} J(\theta) & \mathbf{0} & \mathbf{0} \\ \dot{J}(\theta) & J(\theta) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{5.16}$$

5.4.1 Dynamics compatibility

Since the high-level dynamics is second order whereas the low-level dynamics is third order, dynamics from the two levels are not compatible. In particular, now the low-level control is the muscle activation (see Eq. 5.12), which controls the torque change and cannot affect acceleration on the high-level instantaneously (see Eq. 5.14 and Eq. 5.16), causing $H(\mathbf{x})B(\mathbf{x}) = 0$. However, the change of torque has a predicable effect when applied over time, suggesting that the "instantaneous" $\mathbf{a}$ and B should be replaced with some function that can incorporate temporal predictions. [7] provided an effective way to construct such functions using implicit integration, briefly summarized as follows. First, it discretized the low-level dynamical system (Eq. 5.1) with time step δ:

$$\mathbf{x}(t+\delta) = \mathbf{x}(t) + \delta\left(\mathbf{a}(\mathbf{x}(t)) + B(\mathbf{x}(t))\mathbf{u}(t)\right) \tag{5.17}$$

Then each time step was segmented into n smaller time steps with the same duration δ/n, and Euler integration was performed in each of the n time steps to get

$$\mathbf{x}(t+\delta) = \mathbf{x}(t) + \delta Q(\delta A, n)\left(\mathbf{a}(\mathbf{x}(t)) + B(\mathbf{x}(t))\mathbf{u}(t)\right) \tag{5.18}$$

where $A = \frac{\partial \mathbf{a}}{\partial t}|_{\mathbf{x}=\mathbf{x}_t}$ and

$$Q(X, n) = \frac{1}{n}\left(I + \left(I + \frac{X}{n}\right) + \cdots + \left(I + \frac{X}{n}\right)^{n-1}\right) \tag{5.19}$$

Then the continuous notation was obtained by taking the limit $n \to \infty$, so Eq. 5.18 becomes

$$\mathbf{x}(t+\delta) = \mathbf{x}(t) + \delta Q(\delta A, \infty)\left(\mathbf{a}(\mathbf{x}(t)) + B(\mathbf{x}(t))\mathbf{u}(t)\right) \tag{5.20}$$

Eq. 5.20 differs from Eq. 5.17 only by Q. Therefore, $\mathbf{a}$ and B can be replaced by the "predictive" dynamics $\tilde{\mathbf{a}}$ and $\tilde{B}$, where

$$\begin{aligned} \tilde{\mathbf{a}}(\mathbf{x}) &= Q(\delta A, \infty)\mathbf{a}(\mathbf{x}) \\ \tilde{B}(\mathbf{x}) &= Q(\delta A, \infty)B(\mathbf{x}) \end{aligned} \tag{5.21}$$

Numerically, we have found that $Q(X, n)$ converges as $n > 100$ even for singular X. $\tilde{B}$ now turns into the form

$$\tilde{B} = \begin{bmatrix} \mathbf{0} \\ \mathbf{\Delta} \\ \frac{\mathbf{1}}{\alpha} \end{bmatrix} \tag{5.22}$$

resulting in $H\tilde{B} \neq 0$. The only drawback is the increased computational cost per step. On the other hand, such cost may be offset by the larger time steps that implicit integrations can safely take.

5.4.2 High-level optimization

Ideally, we would like to minimize a cost function that enforces both task performance and energy efficiency. That is,

$$\mathcal{L}(\mathbf{x}, \mathbf{u}) = \int_0^T \left(q(t, \mathbf{x}) + r(t, \mathbf{u})\right) dt \tag{5.23}$$

where $q(t, \mathbf{x})$ and $r(t, \mathbf{u})$ are state-dependent (or task-dependent) cost and control-dependent cost respectively.

As assumed before, $\mathbf{y}$ contains enough information to satisfy $\tilde{q}(t, \mathbf{y}) = q(t, \mathbf{x})$. Thus the state-dependent cost can be fully captured in the high-level. Control-dependent control, on the other hand, may not be fully accounted for from the high-level. Control-dependent cost is usually modeled with a quadratic form

$$r(t, \mathbf{u}) = \frac{\gamma}{2}\mathbf{u}(t)^T\mathbf{u}(t) \tag{5.24}$$

If we can represent low-level control $\mathbf{u}$ using high-level control $\mathbf{v}$ as

$$\mathbf{u}(t) = K(\mathbf{x}, t)\mathbf{v}(t) \tag{5.25}$$

then we can reformulate the control-dependent cost $r(t, \mathbf{u})$ in high-level again in a quadratic form:

$$\tilde{r}(t, \mathbf{v}) = \frac{\gamma}{2}\mathbf{u}(t)^T K(\mathbf{x}, t)^T K(\mathbf{x}, t)\mathbf{u}(t) \tag{5.26}$$

This can be archived by the implicit modeling but not explicit modeling. In particular, applying Eq. 5.15, Eq. 5.16, and Eq. 5.22 to Eq. 5.8, we have

$$P\mathbf{u} = I(\mathbf{y})^{-1}\mathbf{v} \tag{5.27}$$

where $P = J(\theta)\Delta$ and $P \in \mathbb{R}^{n_\mathbf{v}} \times \mathbb{R}^{n_\mathbf{u}}$. Since $n_\mathbf{v} = 3$ whereas $n_\mathbf{u} = 14$, Eq. 5.27 means that only a subspace of the low-level control will affect the task performance and is thus projected to the high-level. Apply singular value decomposition to P and get $P = U\Lambda V^T$. V^T projects the low-level control into task-relevant space (Ω) and task-irrelevant space ($\bar{\Omega}$). Then minimizing $r(t, \mathbf{u})$ becomes equivalent to minimizing the control in Ω and setting $\bar{\Omega}$ into 0. So we get

$$K = (U\tilde{\Lambda})^{-1} I(\mathbf{y})^{-1} \tag{5.28}$$

where $\tilde{\Lambda}$ is the first $n_\mathbf{v}$ columns of Λ.

Although implicit modeling allows the high-level to capture both the passive dynamics and the true cost of the low-level, optimization in the high-level becomes difficult since the high-level is no longer an autonomous dynamical system. Indeed, we apply the iterative LQG (iLQG) method [72], guaranteed to converge for autonomous system, to the high-level, and find that it does yield good solutions. To make the high-level system slightly simpler but still capture the natural dynamics of the low-level, we let the high-level consider only the non-linear skeletal structure of the low-level, while ignore the complex muscle properties. Under the assumption that muscles are strong enough to achieve the torques required by the high-level, the low-level can control the muscle activations to match the high-level controls exactly. After making this modification, the optimization process using iLQG can converge to good solutions after several iterations, which will be presented in numerical simulations.

5.5 Design of low-level controller

In addition to generating energy efficient control $\mathbf{u}$ to satisfy Eq. 5.5, the low-level controller also needs to take into account the constraints of biological movement. First, u_i is constrained to $u_i \in [0, 1]$. Secondly, each joint can only move within a certain range, so $\theta_i \in [\theta_i^{\min}, \theta_i^{\max}]$. In order to punish moving towards the joint limits, we use a polynomial function:

$$g(\theta_i) = (\alpha_i \theta_i + \beta_i)^6 / 100^6 \text{ for } i = 1, 2, ..., 7. \tag{5.29}$$

where

$$\alpha_i = \frac{200}{\theta_i^{\max} - \theta_i^{\min}}, \beta_i = 100 - \frac{200}{\theta_i^{\max} - \theta_i^{\min}} \theta_i^{\max}$$

Here, $\alpha_i\theta_i + \beta_i$ normalizes joint angle $\theta_i \in [\theta_i^{\min}, \theta_i^{\max}]$ into $[-100, 100]$, and 100^6 normalizes $g(\theta_i)$ into $[0, 1]$. Thus $g(\theta_i)$ increase dramatically when the joint angle (θ_i) approaches its joint limit, shown in Fig. 5.3**(b)**.

Now, low-level control $\mathbf{u}$ at each time t can be considered as the solution to the following constrained optimization problem: given $\mathbf{v}$ and $\mathbf{x}$, find $\mathbf{u}$ that minimizes

$$\frac{1}{2}\mathbf{u}^T R\mathbf{u} + \eta\left(\mathbf{a}(\mathbf{x}) + B(\mathbf{x})\mathbf{u}\right)^T \nabla_{\mathbf{x}} \sum_{i=1}^{7} g(\theta_{\mathbf{i}}) \tag{5.31}$$

subject to

$$H(\mathbf{x})B(\mathbf{x})\mathbf{u} = \mathbf{f}(\mathbf{y}) + G(\mathbf{y})\mathbf{v} - H(\mathbf{x})\mathbf{a}(\mathbf{x})$$

$$0 \leq u_i \leq 1 \tag{5.33}$$

where η specifies the weight on moving away from joint limits. This optimization problem can be solved via quadratic programming. Note in implicit modeling, $\mathbf{f}(\mathbf{y}) = H(\mathbf{x})\mathbf{a}(\mathbf{x})$ is guaranteed automatically.

5.6 Numerical simulations

Here we consider three tasks: reaching, orienting and drawing. Our focus is to test the validity of the proposed framework, and also explore the representation of high-level states to account for different task requirements.

5.6.1 Reaching task

The task is to start from some initial position, move and stop the end-effector (fingertip) at a target $\mathbf{p}^*$ defined in Cartesian hand coordinates in a specified time interval T. The cost can be formed in the high-level as follows:

$$\mathcal{L}(\mathbf{y}, \mathbf{v}) = \frac{\gamma}{2}\int_0^T \mathbf{v}(t)^T\mathbf{v}(t)dt + \omega_1\|\mathbf{p}(T) - \mathbf{p}^*\|^2 + \omega_2\|\dot{\mathbf{p}}(T)\|^2 \tag{5.34}$$

where $\mathbf{p}(T)$ and $\dot{\mathbf{p}}(T)$ are the endpoint position and velocity of the end-effector in Cartesian hand coordinates. γ, ω_1 and ω_2 are the relative weights of the terms enforcing energy efficiency, endpoint accuracy and stopping at the target respectively. These three parameters are usually

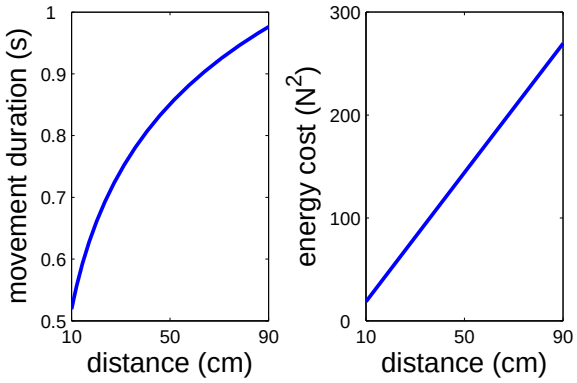


Figure 5.4 Normalization. **(a)**, Total control as a function of movement distance. **(b)**, Time duration as a function of movement distance.

adjusted manually to fit movements with different distance. To automatically scale these parameters, we do the following normalizations. Accuracy and stopping requirements are assumed to be independent of movement distance, therefor their weights (ω_1 and ω_2) are kept constant. The weight of control cost (γ), on the other hand, needs to be scaled according to the distance so that the contribution of energy consumption is relatively the same with respect to that of accuracy and stopping. To get this scaling, we apply an optimal feedback control model [67] to reaching movements with different distance ranging from 10cm to 50cm, and fit the dependency of the total energy consumption ($\Gamma = \frac{1}{2}\int_0^T \mathbf{v}(t)^T\mathbf{v}(t)dt$) on distance ($d \geq 5cm$) and mass (m) on the moving direction as

$$\Gamma = (3.1d - 12.8)m^2 \tag{5.35}$$

Fig. 5.4**(a)** shows Γ as a function of distance. Then we can normalize the control cost by setting $\gamma = \frac{1}{\Gamma}$.

Movement duration T is determined based on Fitt's law [56], which states that the time taken to acquire a visual target should depends on both the movement distance (d) and the accuracy requirement (w) in the form $T = c_1 + c_2 \log_2(\frac{d}{w} + 1)$ where c_1, c_2 are experimentally determined constants. Assuming finishing a movement of 30cm with an accuracy of 1cm takes about 750msec, we set $c_1 = 0$ and $c_2 = 1.5$. Fig. 5.4**(b)** shows T as a function of distance.

We first compare the explicit modeling and implicit modeling on a reaching movement from $\mathbf{p} = [48, -8, -15]^T$ to $\mathbf{p}^* = [21, -21, -30]^T$. In explicit modeling, high-level is considered as a linear system and the optimization problem is solved using LQG [67]. In im-

plicit modeling, high-level is not treated as an autonomous system but accounts for the low-level dynamics. The optimization problem is solved using iLQG. Since iLQG is a local method and may get trapped in local minima, the initial control is provided by LQG which treats the entire arm as a point mass. Fig. 5.5**(a)** shows how the final cost converges over iterations based on the two methods. As we expected, LQG converges within one iteration. However, the cost calculated from the high-level (blue star) does not match that from the low-level (red star). Cost from iLQG, on the other hand, converges after 20 iterations, with the cost calculated from the high-level truthfully captured that from the low-level. Note iLQG guarantees converges only for autonomous dynamical systems, which is why the cost here increases occasionally rather than monotonically decreases. As we can see, the low-level cost after optimization is much smaller from implicit modeling than that from implicit modeling. Fig. 5.5**(b)** further compares the end-effector trajectory during both movement planning and movement execution from both methods. In explicit modeling, the planned trajectory from the high-level (black line) is straight based on the linear dynamics. However, due to the discrepancy of the dynamics from the two levels, this planned trajectory cannot be executed from the low-level, causing big endpoint positional error (green line). The implicit method, in contrast, takes into account the true low-level dynamics and thus planned a curved movement (blue line). This trajectory is tracked exactly by the low-level during execution, leading to much better endpoint accuracy (megenda line). Therefore, the implicit method yields better performance although it takes more computational time. The rest of results are all based on implicit modeling.

Fig. 5.5**(c)** shows the positions and velocities of the end-effector in the Cartesian hand coordinates, as well as the joint angles and velocities in joint coordinates. As we can see, the end-effector velocities in Cartesian coordinates are close to bell-curved profiles, and the end-effector position reaches the target in the end of the movement. Fig. 5.5**(d)** shows the muscle activations during the movement. Note each joint is controlled by two muscles acting on opposite directions (solid lines vs. dashed lines). The two muscle in each muscle pair are activated at different time to either push the joint to move or pull the joint to stop. Fig. 5.5**(e)** and **(f)** show the arm configurations at the beginning and the end of the movement. Note although there are many ways to accomplish the same end-effector position, the arm configuration in **(f)** is a natural posture in the sense that every joint stays in its movement limit. This is archived by including the term $g(\theta)$ on punishing movement approaching joint limits in computing low-level controls (Eq. 5.29 and

Eq. 5.31). Fig. 5.5**(g)** compares the movement of joint 1 and joint 3 before (solid lines) and after (dashed lines) including $g(\theta)$. As we can see, function $g(\theta)$ pushes joint 1 to stay away from its joint limits, and moves joint 3 further to maintain the same end-effector position.

We further test this implicit modeling method with 77 targets evenly distributed in a 3D space surrounding the initial end-effector location. Fig. 5.6**(a)** shows target locations (red circles) and error vector on each target (blue line connecting the end-effector and the corresponding target). Each target location is determined by two parameters: heights and angles on the x-y plane. Fig. 5.6**(b)** and **(c)** show how the error varies over different height (x axis) and angles (represented by different colors of lines) during both planning and execution. As we can see, since the high-level captures the natural plant dynamics, the planned movement is well accomplished by the low-level controller. Also, the normalization of the control cost mentioned before makes it possible to generate movements ranging from 20cm to 50cm with similar accuracy (although error for bigger movement is slightly bigger).

5.6.2 Orienting task

The task is to make a reaching movement and stop at the target, again defined in Cartesian hand coordinates, with a specific palm orientation. One way to solve this problem is to consider a fixed-length pointer attached to the palm with an orientation perpendicular to the palm. The goal of control is to move the two ends of the pointer to two desired positions respectively. The high-level state thus becomes a 18D vector with 9D for each target defined as before. However, this control problem is really hard to solve since now the dynamics in the high-level becomes very complex and iLQG gets trapped in local minia so easily. To make the task slightly simpler, we directly provide the joint angles which will lead to different palm orientations, and define the cost in the high-level as follows

$$\begin{aligned}\mathcal{L}(\mathbf{y},\mathbf{v}) = \frac{\gamma}{2}\int_0^T \mathbf{v}(t)^T\mathbf{v}(t)dt \\ + \omega_1\|\mathbf{p}(T)-\mathbf{p}^*\|^2 + \omega_2\|\dot{\mathbf{p}}(T)\|^2 \\ + \omega_3\|\tilde{\mathbf{p}}(T)-\tilde{\mathbf{p}}^*\|^2 + \omega_4\|\tilde{\dot{\mathbf{p}}}(T)\|^2 \end{aligned} \tag{5.36}$$

where $\mathbf{p}(T)$ and $\dot{\mathbf{p}}(T)$ are the endpoint positions and velocities of the end-effector in Cartesian hand coordinate as defined in reaching tasks. $\tilde{\mathbf{p}}(T)$ and $\tilde{\dot{\mathbf{p}}}(T)$ are the endpoint joint an-

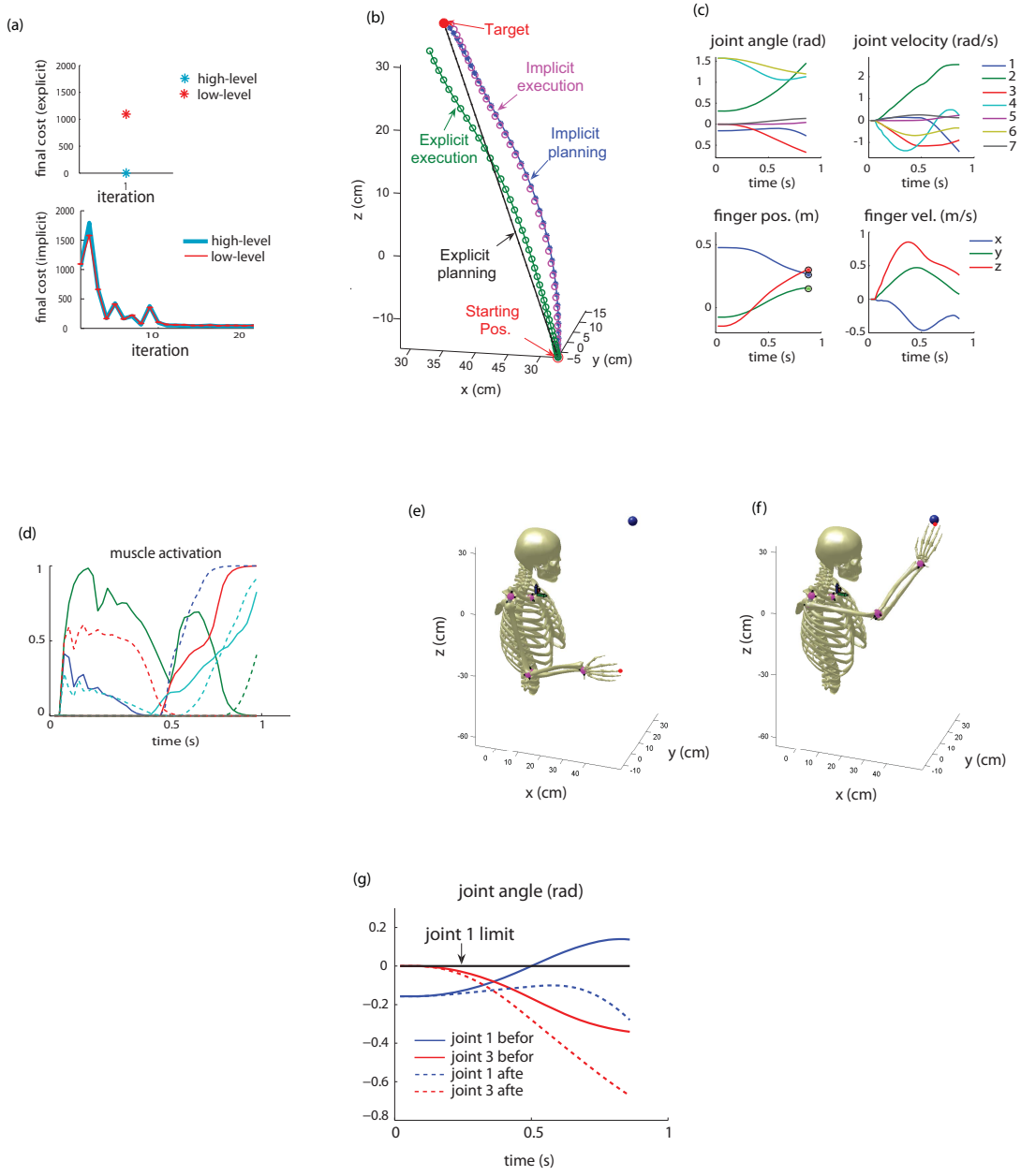


Figure 5.5 Arm movement during reaching. **(a)**, Chang of cost during optimization. **(b)**, Movement trajectory during movement planning and movement execution based on implicit matching or explicit matching. **(c)**, Joint angel, joint velocity, finger position, and finger velocity during a reaching movement. **(d)**, Muscle activations during a reaching movement. **(e)**, Arm configuration at the starting position. **(f)**, Arm configuration at the end position. **(g)**, Comparison of joint moments before and after including cost to punish hitting joint limits.

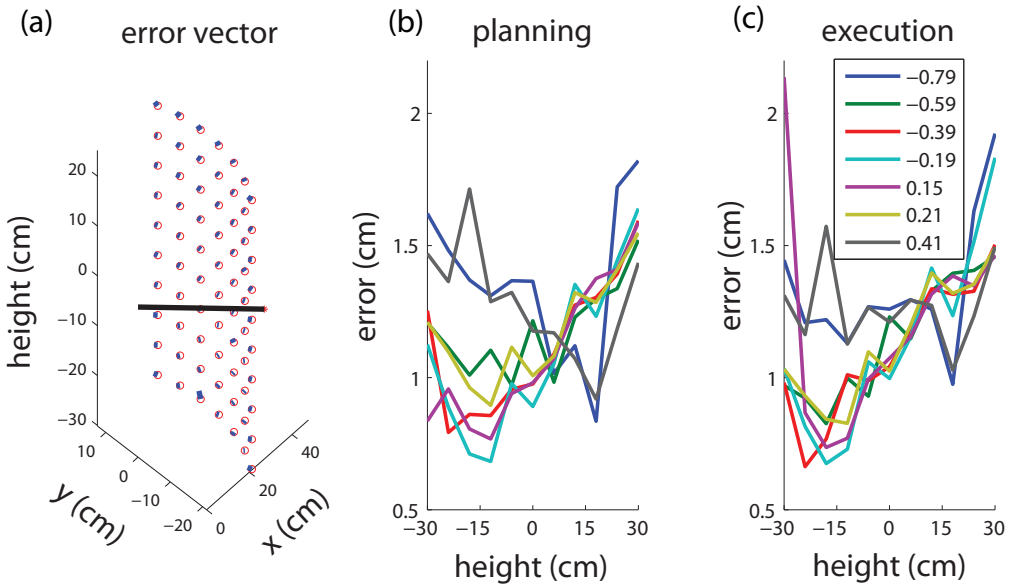


Figure 5.6 Endpoint errors in reaching movements. **(a)**, Target locations and the corresponding error vectors. Each target location is determined by two parameters: height and angle on the x-y coordinate. **(b)**, Endpoint positional errors planned by the high-level. **(c)**, Endpoint positional error executed by the low-level. Lines with different colors represent errors from movements towards targets with different angles.

gles and velocities of specific joints. $\mathbf{p}^*$ and $\tilde{\mathbf{p}}^*$ specify target position and desired joint angles. $\gamma, \omega_1, \omega_2, \omega_3, \omega_4$ are the relative weights of the terms enforcing energy efficiency, endpoint accuracy, stopping at the target, getting specific joint angels, and stopping moving the joints respectively. γ is normalized to account for different moving distance as explained before. $\omega_1, \omega_2, \omega_3, \omega_4$ are adjusted to get the best result.

Fig. 5.7 shows arm configurations in the end of a reaching movement aiming to the same target but with different palm orientations. Fig. 5.7**(a)** is the neutral posture where the wrist is not bended, Fig. 5.7**(b)(c)** are cases requesting for pronation-supination of the palm (movement of joint 5), Fig. 5.7**(d)(e)** are cases requiring flexion-extension of the palm (movement of joint 7). Fig. 5.8 shows how movement of joint 1 is adjusted automatically to accomplish different desired angles of joint 7, while keeping the endpoint end-effector position the same.

5.6.3 Drawing task

The task is to start from some point on a predefined ellipse and track this ellipse as fast as possible. Suppose this ellipse is defined in a 2D plane (e.g., x-z coordinate in the Cartesian hand coordinates with constant value in y coordinate p_y^*), with the two foci $\mathbf{p}_1$ and $\mathbf{p}_2$ and the sum of the distances to these two foci a constant l. The control cost can be defined as in Eq. 5.24, and the state-dependent cost can be formed as follows:

$$q(t, \mathbf{y}) = \omega_1 \|d(t)\|^2 + \omega_2 \frac{\dot{\mathbf{p}}(t) \cdot \mathbf{p}_\perp(t)}{\|\mathbf{p}(t) - \mathbf{o}\|^2} + \omega_3 \|p_y(t) - p_y^*\|^2 \tag{5.37}$$

where $d(t)$ is the tracking error measured as $d(t) = \|\mathbf{p} - \mathbf{p}_1\| + \|\mathbf{p} - \mathbf{p}_2\| - 2l$, $\mathbf{o}$ is the center of the ellipse $\mathbf{o} = \frac{1}{2}(\mathbf{p}_1 + \mathbf{p}_2)$, $\mathbf{p}(t) = [p_x(t), p_z(t)]^T$ and $\dot{\mathbf{p}}(t) = [\dot{p}_x(t), \dot{p}_z(t)]^T$ are the current position and velocity of the end-effector in the x-z coordinate, $\mathbf{p}_\perp(t)$ is a vector on the tangential direction defined as $\mathbf{p}_\perp(t) = [p_z(t), -p_x(t)]^T$, $\frac{\dot{\mathbf{p}}(t) \cdot \mathbf{p}_\perp(t)}{\|\mathbf{p}(t) - \mathbf{o}\|^2}$ computes the angular velocity. $\omega_1, \omega_2, \omega_3$ are the relative weights of the three terms enforcing tracking the ellipse, circling fast, and staying within the 2D plane of the ellipse. As we can see, we do not specify the end-effector position at each time step as in a tracking task, which makes this drawing task even harder.

We apply implicit modeling on a drawing task where the ellipse is simply a circle with the radius of 10cm, and the duration is 6s. Fig. 5.9 **(a)** and **(b)** illustrate the arm configurations before the movement and during the movement. Fig. 5.9 **(c)** show the end-effector trajectories

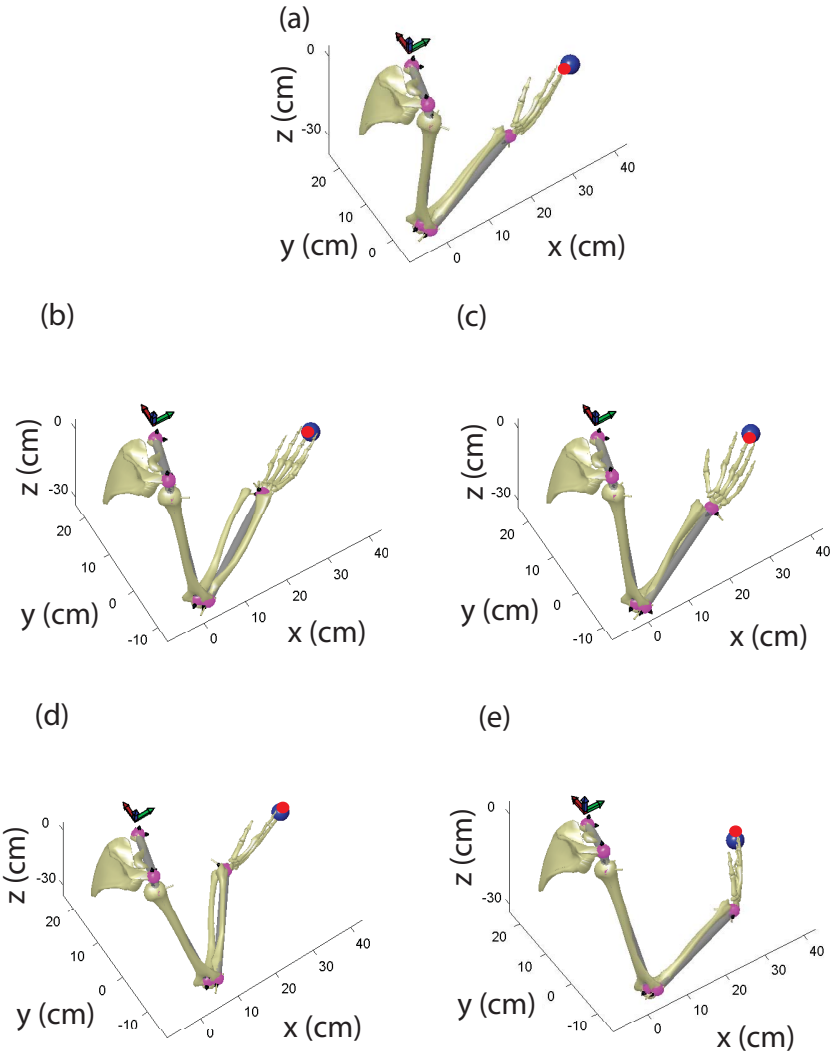


Figure 5.7 Arm configuration with the same end-effector position but different wrist orientations. **(a)**, wrist is not bended. **(b)**, positive pronation-supination of the wrist. **(c)**, negative pronation-supination of the wrist. **(d)**, positive flexion-extension of the wrist. **(e)**, negative flexion-extension of the wrist.

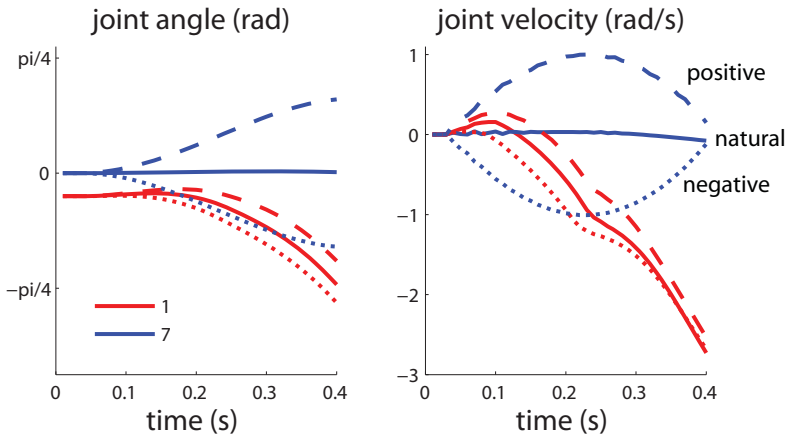


Figure 5.8 Joint angel and joint velocity during a reaching movement with desired palm orientation.

in Cartesian hand coordinates. As we can see, the planned trajectory (light blue line) tracks the desired circle (red line) closely even though the position at each time step is not explicitly specified in the task. This planned trajectory is also well acomplished by the low-level controller (dark blue line). Fig. 5.9 **(d)** shows the joint angles and velocities in joint coordinate, as well as the end-effector positions and velocities in Cartesian hand coordinates. The circular movement involves mainly the movement of joint 2, 4, and 6. Although there is slight oscillation in the beginning, movements become very stable after 0.5s.

5.7 Discussion

A hierarchical control framework is designed for controlling a 7-DOF arm model, which captures most key features of human arm. The high-level feedback controller solves the original control problem but operates on more abstract representations to avoid running into the curse of dimensionality; the low-level feedback controller performs an instantaneous feedback transformation to deal with the details. We show that by allowing the high-level feedback controller to mimic the low-level dynamics, the original optimal control problem can be better captured from the high-level which yields better solutions. To avoid making the high-level too complex to be optimized, the high-level considers only the skeletal structures of the joints rather than the full details of complex muscle properties. Simulation results suggest that taking into

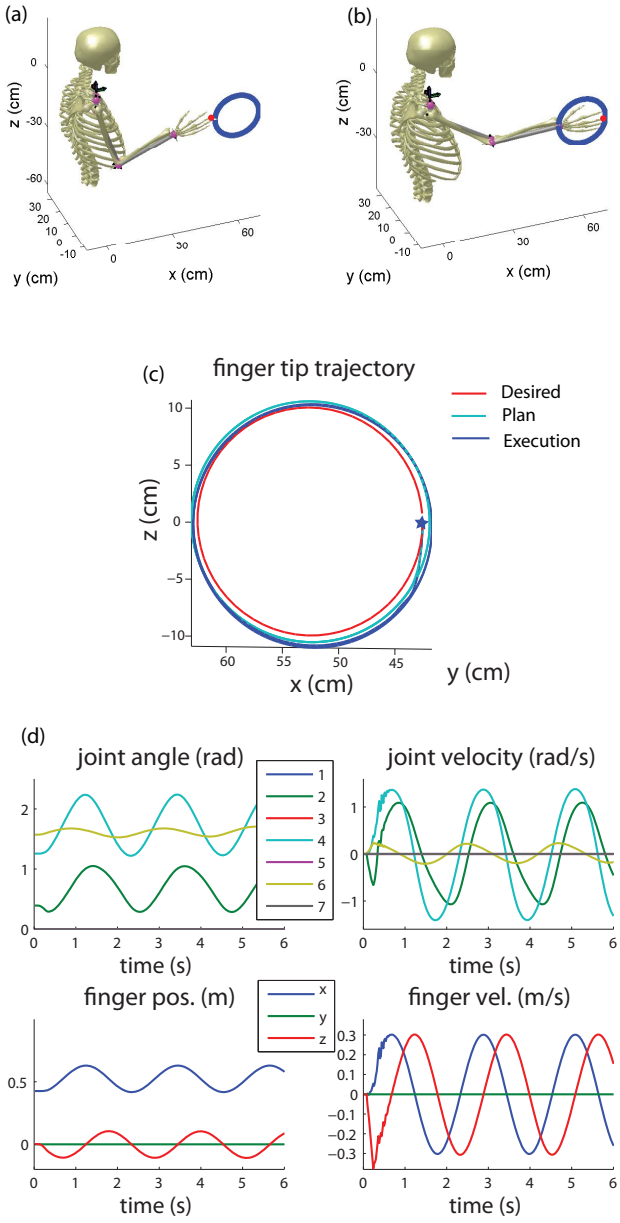


Figure 5.9 Arm movements during drawing. **(a) (b)**, arm configuration at the beginning of movement and during drawing. **(c)**, end-effector trajectory during both planning and execution. **(d)**, position and velocity in both joint coordinates and end-effector Cartesian hand coordinates.

account the low-level dynamics in the high-level generates better results than treating the high-level as a simple linear system. We also provide a scheme to shaping the task-irrelevant space of low-level control to solve the redundancy problem. In particular, by punishing joint movements approaching their limits, the low-level feedback controller generates natural arm configurations to satisfy the task requirements monitored by the high-level. Satisfactory results in tasks such as reaching, orienting and drawing suggest that this hierarchical control framework may not only provide a solution to controlling complex redundant systems, but also shed some light on understanding the neural control of biological movements.

This hierarchical frameworks is more general than other hierarchical schemes aiming to decouple task-level control from details of plant dynamics. For example, the operational space (OS) formulation [110] cannot handle systems other than second order, whereas our hierarchical framework can deal with third-order systems without any modification. This makes it possible for our hierarchical framework to accommodate the true musculoskeletal dynamics of human body. Comparing with feedback linearization (FL) [111, 112] which usually assumes equal numbers of inputs and outputs, our hierarchical framework is able to deal with the mismatch in dimensionality and thus solves the redundancy problem with many more inputs than task-relevant outputs. In addition, instead of augmenting y to handle the mismatch between the high-level and low-level dynamics, our hierarchical controller uses the predictive version $\tilde{\mathbf{a}}$ and $\tilde{B}$ without increasing the dimensionality of the high-level state. Most importantly, OS usually assumes linear dynamics in the high-level and FL tries to linearize the high-level dynamics, both to some extend ignore the complex nonlinearity in the low-level. Our hierarchical framework, in contrast, aims to have the high-level controller exploit the natural plant dynamics and yet operate on reduced dimensionality to better pursue optimality.

Future work includes using more realistic muscle models with non-constant moment arms, examining how control-dependent noise [12, 4] helps to solve the redundancy problem, and testing this framework with more complex tasks.

Acknowledgment

This chapter, in part, has been submitted for publication in *IEEE International Symposium on Adaptive Dynamic Programming and Reinforcement Learning*. The dissertation author

was the primary author of the paper.

Chapter 6

Summary

A main point of this thesis is that during goal-directed movement, the CNS exploits every opportunity to best archive the goal, instead of depending on a fixed control strategy. Such flexibility has been largely ignored by traditional theories emphasizing movement geometry and servo control mechinisms, but lies at the core of optimal feedback control.

The first form of flexibility we studied was the sensitivity of motor planning to task goals, presented in Chapter 3. Our point of departure was the little-known fact that corrections for target perturbations introduced late in a reaching movement are incomplete. We showed that this was not simply attributable to the lack of time, in contradiction with alternative models and, somewhat paradoxically, in agreement with our optimal feedback control model. Analysis of optimal feedback gains reveals that the effect was partly attributable to a previously unknown trade-off between stability and accuracy. This yielded a testable prediction: if stability requirements were decreased, then accuracy should increase. We confirmed the prediction experimentally in three-dimensional obstacle avoidance and interception tasks in which subjects hit a robotic target with programmable impedance. In additional agreement with the theory, we found that subjects exploited every opportunity for increased performance. These results suggest that the remarkable flexibility of motor behavior arises from sensorimotor control laws optimized for composite cost functions.

The second form of flexibility we studied was motor system's adaptability in a changing environment, presented in Chapter 4. By experimentally changing the perturbation time and dynamic environments during a reaching movement, we showed that prediction of perturbations

was used in on-line corrections only if it can lead to better endpoint positional accuracy and less energy consumptions. These results suggest that motor learning involves both the statistical formulation of internal models to predict how the environment changes, and the flexible use of internal models in adjusting motor commands to best achieve the behavior goal. This view challenges the implicit assumption in the literature that the output of internal models is in one-to-one correspondence with changes in motor behavior. The flexible use of internal models in generating motor commands was largely explained within the framework of optimal feedback control. This required some novel extensions to the framework. The extensions have to do with robustness (as in high-order energy costs) as well as a preference for control strategies that tend to be successful in everyday life outside the lab.

The last form of flexibility we focused on arises from the redundant musculoskeletal structure of human body. In Chapter 5 we applied a hierarchical control framework to a more realistic arm model with 7 degrees of freedom and 14 muscles whose dynamics resemble real muscle dynamics. By designing the high-level controller to account for the complex low-level dynamics but operate on more abstract representations, we were able to solve the optimal control problem with respect to the non-linear musculoskeletal dynamics without running into the curse of dimensionality. Meanwhile, the low-level controller solved the redundancy problem by considering both energy efficiency and biological constraints during optimization. Satisfactory results in tasks such as reaching, drawing and orienting suggest that this hierarchical control framework may provide us some insight on how the brain controls human body which is a complex redundant system.

Bibliography

[1] S. Scott, "Optimal feedback control and the neural basis of volitional motor control," *Nature Neuroscience*, vol. 5, pp. 534–546, 2004.

[2] K. Doya, "Reinforcement learning in continuous time and space," *Neural Computation*, vol. 12, pp. 219–245, 2000.

[3] P. Morasso, "Spatial control of arm movements," *Exp Brain Res*, vol. 42, pp. 223–227, 1981.

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

[5] D. Wolpert, "Computational approaches to motor control," *Trends in Cognitive Science*, vol. 1, no. 6, pp. 209–216, 1997.

[6] E. Todorov, *Optimal Control Theory*, ch. 12. Bayesian Brain, Doya, K. (Ed), Cambridge, MA: MIT Press, 2006.

[7] E. Todorov, W. Li, and X. Pan, "From task parameters to motor synergies: A hierarchical framework for approximately-optimal feedback control of redundant manipulators," *Journal of Robotic Systems*, vol. 22, pp. 691–710, 2005.

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

[9] K. Kording and D. Wolpert, "The loss function of sensorimotor learning," *Proceedings of the National Academy of Sciences*, vol. 101, pp. 9839–9842, 2004.

[10] 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.

[11] Y. Uno, M. Kawato, and R. Suzuki, "Formation and control of optimal trajectory in human multijoint arm movement: Minimum torque-change model," *Biological Cybernetics*, vol. 61, pp. 89–101, 1989.

[12] C. Harris and D. Wolpert, "Signal-dependent noise determines motor planning," *Nature*, vol. 394, pp. 780–784, 1998.

[13] A. Feldman, "Once more on the equilibrium-point hypothesis (lamda-model) for motor control," *Journal of Motor Behavior*, vol. 18, pp. 17–54, 1986.

[14] A. Feldman and M. Levin, "The origin and use of positional frames of reference in motor control," *Behavioral and Brain Sciences*, vol. 18, no. 4, pp. 723–744, 1995.

[15] E. Bizzi, N. Hogan, F. Mussa-Ivaldi, and S. Giszter, "Does the nervous system use equilibrium-point control to guide single and multiple joint movements?," *Behavioral and Brain Sciences*, vol. 15, pp. 603–613, 1992.

[16] E. Todorov and M. Jordan, "A minimal intervention principle for coordinated movement," in *Advances in Neural Information Processing Systems 15* (S. Becker, S. Thrun, and K. Obermayer, eds.), pp. 27–34, Cambridge, MA: MIT Press, 2002.

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

[18] N. Bernstein, *The Coordination and Regulation of Movements*. Pergamon Press, 1967.

[19] K. Cole and J. Abbs, "Kinematic and electromyographic responses to perturbation of a rapid grasp," *J Neurophysiol*, vol. 57, no. 5, pp. 1498–510, 1987.

[20] J. Scholz and G. Schoner, "The uncontrolled manifold concept: identifying control variables for a functional task," *Exp Brain Res*, vol. 126, no. 3, pp. 289–306, 1999.

[21] D. Domkin, J. Laczko, S. Jaric, H. Johansson, and M. Latash, "Structure of joint variability in bimanual pointing tasks," *Exp Brain Res*, vol. 143, pp. 11–23, 2002.

[22] W. Nelson, "Physical principles for economies of skilled movements," *Biological Cybernetics*, vol. 46, pp. 135–147, 1983.

[23] F. Anderson and M. Pandy, "Dynamic optimization of human walking," *J Biomech.Eng*, vol. 123, no. 5, pp. 381–390, 2001.

[24] E. Todorov, "Cosine tuning minimizes motor errors," *Neural Computation*, vol. 14, no. 6, pp. 1233–1260, 2002.

[25] G. Sutton and K. Sykes, "The variation of hand tremor with force in healthy subjects," *Journal of Physiology*, vol. 191(3), pp. 699–711, 1967.

[26] R. Schmidt, H. Zelaznik, B. Hawkins, J. Frank, and J. Quinn, "Motor-output variability: a theory for the accuracy of rapid notor acts," *Psychol Rev*, vol. 86, no. 5, pp. 415–451, 1979.

[27] A. Hamilton, K. Jones, and D. Wolpert, "The scaling of motor noise with muscle strength and motor unit number in humans," *Exp Brain Res*, vol. 157, pp. 417–430, 2004.

[28] Z. Ghahramani, D. Wolpert, and M. Jordan, "Generalization to local remappings of the visuomotor coordinate transformation," *Journal of Neuroscience*, vol. 16, pp. 7085–7096, 1996.

[29] Z. Ghahramani and D. Wolpert, "Modular decomposition in visuomotor learning.," *Nature*, vol. 386, pp. 392–395, 1997.

[30] S. Sober and P. Sabes, "Flexible strategies for sensory integration during motor planning," *Nature Neuroscience*, vol. 8, pp. 490–497, 2005.

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

[32] O. Donchin, J. Francis, and R. Shadmehr, "Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: Theory and experiments in human motor control.," *J Neurosci*, vol. 23, no. 27, pp. 9032–9045, 2003.

[33] D. Wolpert, Z. Gharahmani, and M. Jordan, "An internal model for sensorimotor integration," *Science*, vol. 269, pp. 1880–1882, 1995.

[34] M. Kawato, "Internal models for motor control and trajectory planning," *Curr Opin Neurobiol*, vol. 9, no. 6, pp. 718–27, 1999.

[35] M. Ernst and M. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, pp. 429–433, 2002.

[36] K. Kording, S. Ku, and D. Wolpert, "Bayesian integration in force estimation," *J.Neurophysiology*, vol. 92, pp. 3161–3165, 2004.

[37] K. Kording and D. Wolpert, "Bayesian decision theory in sensorimotor control," *Trends in Cognitive Sciences*, vol. 10, no. 7, pp. 319–326, 2006.

[38] K. Kording, J. Tenenbaum, and R. Shadmehr, "The dynamics of memory as a consequence of optimal adaptation to a changing body," *Nature Neuroscience*, vol. 10, pp. 779–786, 2007.

[39] N. Bernstein, "Dexterity and its development," in *Dexterity and its development* (M. Latash and M. Turvey, eds.), Lawrence Erlbaum, 1996.

[40] G. Loeb, I. Brown, and E. Cheng, "A hierarchical foundation for models of sensorimotor control," *Exp Brain Res*, vol. 126, no. 1, pp. 1–18, 1999.

[41] J. Kalaska, L. Sergio, and P. Cisek, "Cortical control of whole-arm motor tasks," in *Sensory guidance of movement: Novartis Foundation Symposium* (M. Glickstein, ed.), Chichester, UK: John Wiley and Sons, 1998.

[42] C. Sherrington, *The Integrative Action of the Nervous System*. New haven: Yale University Press, 1947.

[43] A. Georgopoulos, J. Kalaska, R. Caminiti, and J. Massey, "On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex," *J.Neurosci.*, vol. 2, no. 11, pp. 1527–1537, 1982.

[44] M. Desmurget, C. Epstein, R. Turner, C. Prablanc, G. Alexander, and S. Grafton, "Role of the posterior parietal cortex in updating reaching movements to a visual target," *Nat.Neurosci.*, vol. 2, no. 6, pp. 563–567, 1999.

[45] R. Andersen, L. Snyder, D. Bradley, and J. Xing, "Multimodal representation of space in the posterior parietal cortex and its use in planning movements," *Annual Review of Neuroscience*, vol. 20, pp. 303–330, 1997.

[46] J. Fuster, *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of The. Frontal Lobe*. Lippincott-Raven Press, 1997.

[47] A. Pasupathy and E. Miller, "Different time courses of learning-related activity in the prefrontal cortex and striatum," *Nature*, vol. 433, pp. 873–876, 2005.

[48] M. Smith and R. Shadmehr, "Ntact ability to learn internal models of arm dynamics in huntington's disease but not cerebellar degeneration," *Journal of Neurophysiology*, vol. 93, no. 5, pp. 2809–2821, 2005.

[49] T. Vilis and J. Hore, "Effects of changes in mechanical state of limb on cerebellar intention tremor," *Journal of Neurophysiology*, vol. 40, no. 5, pp. 1214–1224, 1977.

[50] M. Demirci, S. Grill, L. McShane, and M. Hallett, "A mismatch between kinesthetic and visual perception in parkinson's disease," *Ann Neurol*, vol. 41, pp. 781–788, 1997.

[51] M. Ito, M. Sakurai, and P. Tongroach, "Climbing fibre induced depression of both mossy fibre responsiveness and glutamate sensitivity of cerebellar purkinje cells," *Journal of Physiology*, vol. 324, pp. 113–134, 1982.

[52] A. Graybiel, "The basal ganglia: Learning new tricks and loving it," *Curr Opin Neurobiol*, vol. 15, pp. 638–644, 2005.

[53] C. Fiorillo, P. Tobler, and W. Schultz, "Discrete coding of reward probability and uncertainty by dopamine neurons," *Science*, vol. 299, pp. 1898–1902, 2003.

[54] H. Nakahara, H. Itoh, R. Kawagoe, Y. Takikawa, and O. Hikosaka, "Dopamine neurons represent context-dependent reward prediction error," *Neuron*, vol. 41, pp. 269–280, 2004.

[55] P. N. Tobler, C. D. Fiorillo, and W. Schultz, "Adaptive coding of reward value by dopamine neurons," *Scienc*, vol. 307, pp. 1642–1645, 2005.

[56] P. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *Journal of Motor Behavior*, vol. 21, pp. 5–19, 1954.

[57] J. Medina and S. Lisberger, "Ariation, signal, and noise in cerebellar SensoryŰMotor processing for smooth-pursuit eye movements," *Journal of Neuroscience*, vol. 27, no. 25, pp. 6832–6842, 2007.

[58] H. Gomi and M. Kawato, "Equilibrium-point control hypothesis examined by measured arm stiffness during multijoint movement," *Nature*, vol. 272, pp. 117–120, 1996.

[59] R. Shadmehr and S. Wise, *Omputational Neurobiology of Reaching and Pointing: A Foundation for Motor Learning*. MIT Press, Cambridge MA, 2005.

[60] T. Flash and E. Henis, "Arm trajectory modification during reaching towards visual targets," *Journal of Cognitive Neuroscience*, vol. 3, pp. 220–230, 1991.

[61] B. Rohrer and N. Hogan, "Avoiding spurious submovement decompositions: A globally optimal algorithm," *Biological Cybernetics*, vol. 89, no. 3, pp. 190–199, 1989.

[62] E. Kandel, J. Schwartz, and T. Jessell, *Principles of Neural Science*. McGraw-Hill Education, 2000.

[63] S. Blakemore, D. Wolpert, and C. Frith, "Central cancellation of self-produced tickle sensation," *Nature Neuroscience*, vol. 1, pp. 635 – 640, 1998.

[64] R. Kalman, "When is a linear control system optimal?," *Trans AMSE J Basic Eng, Ser D*, vol. 86, pp. 51–60, 1964.

[65] M. Davis and R. Vinter, *Stochastic Modelling and Control*. Chapman and Hall, London, 1985.

[66] E. Todorov, "Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system," *Neural Comp*, vol. 17, no. 5, pp. 1084–1108, 2005.

[67] E. Todorov and W. Li, "A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems," *American Control Conference*, 2005.

[68] H. Chen-Harris, W. Joiner, V. Ethier, D. Zee, and R. Shadmehr, "Adaptive control of saccades via internal feedback," *Journal of Neuroscience*, vol. 28, pp. 2804–2813, 2008.

[69] H. Kushner and P. Dupuis, *Numerical Methods for Stochastic Optimal Control Problems in Continuous Time*. New York: Springer, 2 ed., 2001.

[70] D. Bertsekas, *Dynamic Programming and Optimal Control (2nd Ed)*. Bellmont, MA: Athena Scientific, 2001.

[71] D. Liu and E. Todorov, "Evidence for the flexible sensorimotor strategies predicted by optimal feedback control," *Journal of Neuroscience*, vol. 27, pp. 9354–9368, 2007.

[72] E. Todorov and W. Li, "A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems," in *American Control Conference*, 2005.

[73] D. Pelisson, C. Prablanc, M. Goodale, and M. Jeannerod, "Visual control of reaching movements without vision of the limb. ii. evidence of fast unconscious processes correcting the trajectory of the hand to the final position of a double-step stimulus," *Exp Brain Res*, vol. 62, no. 2, pp. 303–11, 1986.

[74] C. Prablanc and O. Martin, "Automatic control during hand reaching at undetected two-dimensional target displacements," *Journal of Neurophysiology*, vol. 67, pp. 455–469, 1992.

[75] R. Scheidt, B. Jonathan, and A. Ferdinando, "Learning to move amid uncertainty," *Journal of Neurophysiology*, vol. 86, pp. 971–985, 2001.

[76] S. Cheng and P. Sabes, "Modeling sensorimotor learning with linear dynamical systems," *Neurcal Comp*, vol. 18, pp. 760–793, 2006.

[77] J. Diedrichsen, Y. Hashambhoy, T. Rane, and R. Shadmehr, "Neural correlates of reach errors," *J Neurosci*, vol. 25, pp. 9919–9931, 2005.

[78] J. Trommershäuser, S. Gepshtein, L. Maloney, M. Landy, and M. Banks, "Optimal compensation for changes in task-relevant movement variability," *Journal of Neuroscience*, vol. 25, no. 31, pp. 7169–7178, 2005.

[79] K. Kording, S. Ku, and D. Wolpert, "Bayesian integration in force estimation," *Journal of Neurophysiology*, vol. 92, pp. 3161–316, 2004].

[80] J. Saunders and D. Knill, "Visual feedback control of hand movements," *Journal of Neuroscience*, pp. 3223–3234, 2004.

[81] M. Smith, A. Ghazizadeh, and R. Shadmehr, "Interacting adaptive processes with different timescales underlie short-term motor learning," *PLoS Biology*, vol. 4, p. e179, 2006.

[82] T. Behrens, M. Woolrich, M. Walton, and M. Rushworth, "Learning the value of information in an uncertain world," *Nature Neuroscience*, vol. 10, pp. 1214 – 1221, 2007.

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

[84] D. Meyer, R. Abrams, S. Kornblum, C. Wright, and J. Smith, "Optimality in human motor performance: Ideal control of rapid aimed movements," *Psychological Review*, vol. 95, pp. 340–370, 1988.

[85] G. Loeb, W. Levine, and J. He, "Understanding sensorimotor feedback through optimal control," *Cold Spring Harb.Symp.Quant.Biol.*, vol. 55, pp. 791–803, 1990.

[86] B. Hoff, *A computational description of the organization of human reaching and prehension*. Ph.D. Thesis, University of Southern California, 1992.

[87] A. Kuo, "An optimal control model for analyzing human postural balance," *IEEE Transactions on Biomedical Engineering*, vol. 42, pp. 87–101, 1995.

[88] M. Jeannerod, *The Neural and Behavioural Organization of Goal-Directed Movements.* Oxford University Press, 1988.

[89] M. Desmurget and S. Grafton, "Forward modeling allows feedback control for fast reaching movements," *Trends Cogn Sci*, vol. 4, no. 11, pp. 423–431, 2000.

[90] G. Hinton, "Parallel computations for controlling an arm," *Journal of Motor Behavior*, vol. 16, no. 2, pp. 171–194, 1984.

[91] B. Hoff and M. Arbib, "Models of trajectory formation and temporal interaction of reach and grasp," *J Mot. Behav.*, vol. 25, no. 3, pp. 175–192, 1993.

[92] E. Torres and D. Zipser, "Reaching to grasp with a multi-jointed arm. i. computational model," *Journal of Neurophysiology*, vol. 88, no. 5, pp. 2355–2367, 2002.

[93] E. Komilis, D. Pelisson, and C. Prablanc, "Error processing in pointing at randmoly feedback-induced double-step stimuli," *Journal of Motor Behavior*, vol. 25, no. 4, pp. 299–308, 1993.

[94] F. Popescu and W. Rymer, "End points of planar reaching movements are disrupted by small force pulses: An evaluation of the hypothesis of equifinality," *Journal of Neurophysiology*, vol. 84, pp. 2670–2679, 2000.

[95] T. Nicols and J. Houk, "Improvement in linearity and regulations of stiffness that result from actions of stretch reflex.," *Journal of Neurophysiology*, vol. 39, pp. 119–142, 1976.

[96] W. Li and E. Todorov, "Iterative optimal control and estimation design for nonlinear stochastic systems," *IEEE Control and Decision Conference*, 2006.

[97] R. Van Beers, A. Sittig, and J. Van der Gon, "Integration of proprioceptive and visual position-information: An experimentally supported model," *Journal of Neurophysiology*, vol. 81, pp. 1355–1364, 1999.

[98] C. Harris, "Does saccadic undershoot minimize saccadic flight-time? a monte-carlo study," *Vision Research*, vol. 35, pp. 691–701, 1995.

[99] D. Hoffman and P. Strick, "Step-tracking movements of the wrist. IV. muscle activity associated with movements in different directions," *Journal of Neurophysiology*, vol. 81, pp. 319–333, 1999.

[100] M. Haruno and D. Wolpert, "Optimal control of redundant muscles in step-tracking wrist movements," *Journal of Neurophysiology*, vol. 94, pp. 4244–4255, 2005.

[101] J. Lackner and P. Dizio, "Rapid adaptation to coriolis-force perturbations of arm trajectory," *Journal of Neurophysiology*, vol. 72, no. 1, pp. 299–313, 1994.

[102] M. Hinder and T. Milner, "The case for an internal dynamics model versus equilibrium point control in human movement," *Journal of Physiology*, vol. 549, pp. 953–963, 2003.

[103] D. Rosenbaum, R. Meulenbroek, J. Vaughan, and C. Jansen, "Coordination of reaching and grasping by capitalizing on obstacle avoidance and other constraints," *Exp Brain Res*, vol. 128, pp. 92–100, 1999.

[104] A. Hamilton and D. Wolpert, "Controlling the statistics of action: obstacle avoidance," *J Neurophysiol.*, vol. 87, no. 5, pp. 2434–2440, 2002.

[105] P. Gribble, L. Mullin, N. Cothros, and A. Mattar, "Role of cocontraction in arm movement accuracy," *Journal of Neurophysiology*, vol. 89, pp. 2396–2405, 2003.

[106] M. Pandy, "Computer modeling and simulation of human movement," *Annu. Rev. Biomed. Eng.*, vol. 3, pp. 245–273, 2001.

[107] R. Bellman, *Dynamic Programming*. Princeton, NJ.: Princeton University Press, 1957.

[108] W. Li and E. Todorov, "Iterative linear-quadratic regulator design for nonlinear biological movement systems," in *1st International Conference on Informatics in Control, Automation and Robotics*, 2004.

[109] I. Brown, S. Scott, and G. Loeb, "Mechanics of feline soleus: II. design and validation of a mathematical model," *Journal of Muscle Research and Cell Motility*, vol. 17, pp. 219–223, 1996.

[110] O. Khatib, "A unified approach to motion and force control of robotic manipulators: The operational space formulation," *IEEE Journal of Robotics and Automation*, vol. RA-3, no. 1, pp. 43–53, 1987.

[111] A. Isidori, *Nonlinear Control Systems*. London: Springer-Verlag, 1995.

[112] H. Khalil, *Nonlinear Systems*. New Jersey: Prentice-Hall, 2002.