NAIST-IS-DD0261203

Doctoral Dissertation

Real-time constraints to learning and control of voluntary movement

Fredrik Bissmarck

March 5, 2008

Department of Bioinformatics and Genomics
Graduate School of Information Science
Nara Institute of Science and Technology

A Doctoral Dissertation
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of SCIENCE

Fredrik Bissmarck

Thesis Committee:

Professor Shin Ishii	(Supervisor)
Professor Kenji Doya	(Co-supervisor)
Professor Mitsuo Kawato	(Co-supervisor)
Professor Kotaro Minato	(Co-supervisor)

Real-time constraints to learning and control of voluntary movement*

Fredrik Bissmarck

Abstract

The plasticity and computational capacity of the human cerebral cortex offer great potential for planning, learning and execution of movements. Indeed, a large part of the cortex is recruited for motor control and learning. However, a limitation is the long latencies of feedback from sensors and actuators of the peripheral nervous system - up to 100's of milliseconds. This constraint imposes a challenge to utilize the cerebral cortex for real-time motor control.

This thesis seeks to elucidate the real-time constraints of cortical feedback loops for motor control. A broader aim of our study is the understanding of long-term learning of sequential, manual movement. We investigate how a series of planned movements are gradually integrated into a fast, skillful, single movement, and how the recruitment of sensory feedback may be altered through stages of learning. We present two different studies on this theme. In the first, we take a computational approach. Proposing a framework with analogies of the basal ganglia-thalamocortical system, we address the problem of combining multiple feedback modalities of different latencies to learn joint torque controlled arm movements. In the second, we take an experimental approach, and study the long-term alteration of gaze strategies in a manual task.

We first review related work and important concepts: theories of optimal motor control, including reinforcement learning (Chapter 2), and then empirical

*Doctoral Dissertation, Department of Bioinformatics and Genomics, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0261203, March 5, 2008.

psychology and neurophysiology of visuomotor coordination and learning of goal-directed, sequential movements (Chapter 3). Then, we present the computational study (Chapter 4). We propose a general framework for combining and learning modalities with different latencies, based on the actor-critic algorithm. In a first simple implementation of a somatosensory reaching task, we assert our hypotheses that, given identical modules of different feedback latencies, 1) performance is limited by the latency of the faster module alone, and 2) that the faster module becomes dominant over control. In a second implementation, we examined an example of visuomotor sequence learning, where a plastic, faster somatosensory module interacts with a preacquired, slower visual module. Here we find that the somatosensory module acquires an independent control policy with better performance than the visual module. The visual module displays differential roles; in the early learning stage, it acts as a guide for the somatosensory module, and in the late learning stage, it acts as a safeguard against perturbations.

In the following chapter (Chapter 5), we present the experimental study. We first introduce our paradigm to investigate the long term behavioural change in a stereotype, sequential button pressing task. We present a Bayesian model of dynamic updating of spatial representation, with the potential to explain gaze behaviour for manual tasks. We then report our findings of changes of gaze: in early learning, subjects fixate each target button, but as the manual execution speeds up, subjects fixate strategic points, inclined towards center-of-mass of clusters of targets. We also provide evidence that the Bayesian model can explain gaze-dependence of manual accuracy.

Overall, our findings contribute to the understanding of the real-time limitations of cortical feedback systems, and what consequences it may have for visuomotor feedback control and learning, in particular for 1) reward-based basal ganglia learning functions, and 2) gaze strategies for manual skills.

Keywords:

Motor control, reinforcement learning, basal ganglia, gaze behaviour, eye-hand coordination, motor sequence learning

随意運動の学習と制御における実時間の制約*

ビスマルク・フレデリック

内容梗概

身体運動の計画・遂行および学習は、ヒト大脳皮質の計算能力と可塑性によって可能になる。実際、大脳の多くの部位は運動制御とその学習に使われている。運動制御・学習において問題となるのは、筋肉や感覚受容器から生じる数百ミリ秒にもなる感覚フィードバックにおける遅れである。実時間での運動制御において、この時間遅れを大脳皮質はどのように解いているのかは容易い問題ではない。

本論文では、運動制御において大脳フィードバックに課された実時間制限がどのように解かれているかを明らかにしようと試みる。さらに、手の系列運動が長期にわたりどのように学習されるかを議論する。特に、計画された一連の運動がどのように途切れの無い熟達した運動に統合されるのか、また、学習が進むにつれて感覚フィードバックが運動学習における使われ方がどのように変化するのか、といった問題に取り組む。これらの問題に対して、二つの異なる研究を示す。まず始めの研究では、計算論的アプローチをとり、大脳基底核視床系からヒントを得たモデルを提唱する。そのモデルにおいて、上腕運動のトルク制御する際、複数のモダリティに由来する異なる潜時の感覚フィードバックをどのように用いるかを議論する。二つ目の研究では実験的なアプローチをとり、手の系列運動を学習する際、注視のパターンが長期的にどのように変化したかを調べた行動実験について報告する。

本論文の構成についてまとめる。まず関連する先行研究および概念について解説する。第2章では、強化学習といった、最適制御理論に基づく運動制御の理論を概説し、第3章では合目的な系列運動や視覚運動協調において知られている心理学・神経生理学知見をまとめる。その上で、計算論の研究について述べる。強化学習でのアクター・クリティックと呼ばれるアルゴリズムに基づき、異なる

*奈良先端科学技術大学院大学 情報科学研究科 情報生命科学専攻 博士論文, NAIST-IS-DD0261203, 2008年3月5日.

遅れ時間を持つ複数のモダリティからのフィードバックをどのように統合しまた学習するかを理解するための一般的な枠組みを提案する。この枠組みをまず到達運動に適用し、遅れ時間のみが異なる複数のフィードバック系を考える。そこで最も遅れが短いモジュールのみが使われること、そしてそれが到達運動の達成度を決定することを示す。次にこの枠組みを視覚運動系列学習の例に適用する。そこでは、可塑性のない遅い視覚モジュールが可塑性のある体性感覚モジュールと相互作用すると仮定する。シミュレーションから、体性感覚モジュールは視覚モジュールとは独立のより達成度の高い制御ポリシーを獲得することを示す。視覚モジュールは学習が進むにつれて異なる役割を担う。学習の初期においては体性感覚モジュールの学習のお手本となり、学習が進んだあとでは擾乱に対して安定性を保証する。

第5章では、行動実験について解説する。まず、順序を固定したボタン押し系列学習において、長期にわたる行動変化を調べた実験パラダイムを紹介する。その運動学習における眼球運動の振る舞いを説明するために、ベイズ理論に基づいた、動的に空間表現を更新するモデルを提案する。学習初期においては注視方向は次に押そうとするボタンに向かうが、学習が進み手先の運動が速くなるにつれて、次に押そうとするボタンそのものではなく、複数のボタンが集団を成している中心に注視方向は向かうようになる。ベイズモデルは手運動の正確さが注視依存であることを説明する。

ここで提案した計算論的研究により、感覚フィードバック制御の限界を定量的に明らかにすることができる。加えて、どのモダリティを用いるかを明示的に指定することなく、最適な行動を与えるフィードバックへの重みを強化することにより、モダリティ間の統合を柔軟に行う描像を与えた。次に解説した実験的研究により、視覚は熟達した運動には必要だが、最適フィードバックのために注視方向を移す速度には限界があることを示した。この注視運動の速度限界のため、注視計画は手運動の熟達により変化する。

キーワード

運動制御、大脳基底核、強化学習、注視行動、系列学習

Acknowledgements

First and foremost, it has been a great privilege to work with my supervisor Kenji Doya. I have learned so much from Kenji about the computational and neuro-sciences. His thorough feedback on all aspects of my work has helped me to become a better scientist. Furthermore, Kenji also taught me the value of diligence as well as the virtue of expressing science effectively and precisely. I also thank my formal supervisor Shin Ishii for helping out with administrative issues.

It has also been a great pleasure to collaborate with David Franklin, Hiroyuki Nakahara, and Okihide Hikosaka. From David I learned a lot about practical issues of experimental neuroscience. Hiroyuki was always helpful and encouraging on revision of manuscripts.

Throughout the programme, I have on many occasions received valuable feedback and encouragement from Mitsuo Kawato, Erhan Oztop, Ganesh Gowrishankar, Hirokazu Tanaka, Nicholas Schweighofer, Yukiyasu Kamitani, Jun Morimoto and Daniel Callan. They all contributed to improve on ideas, results and manuscripts. The technical support group always provided excellent support without delay: Noboru Nushi, Mitsutoshi Uchida, Kojiro Fujii and Satoshi Tada. Norikazu Sugimoto also provided invaluable help with technical issues on numerous occasions. I am also grateful to a number of people at ATR and NAIST for helping out with various practical issues, particularly: Sae Franklin, Satomi Higuchi, Mami Inada, Saori Tanaka, Hirofumi Suzuyama, Yasunobu Igarashi, Tomokazu Doi, Takamitsu Matsubara, Gary Liaw, Akiko Callan and Akiko Nonaka.

I am grateful to the Japanese Ministry of Education (Monbusho) for funding most of my PhD programme.

I thank my family - Boel, Georg & Alice, Adam and Miyuki for the support, and their patience with me during this time.

This work is dedicated to the memory of my uncle Stefan, to whom I am most grateful. He taught me, with the innocent child ’ s eye, about the pleasure of finding things out.

Contents

List of publications

Journal articles

F. Bissmarck, H. Nakahara, K. Doya, and O. Hikosaka. (in press) Combining modalities with different latencies for optimal motor control. *Journal of Cognitive Neuroscience.*

F. Bissmarck, D. W. Franklin, and K. Doya. (in review). Altered gaze strategies in sequential hand movement.

Conference papers

F. Bissmarck, D. W. Franklin, and K. Doya. Selective saccades in sequential hand movements. *IEICE Technical Report*, 105(34), 1-5, 2005.

F. Bissmarck, H. Nakahara, K. Doya, and O. Hikosaka. Responding to modalities with different latencies. *Advances in Neural Information Processing Systems (NIPS)*, 17:169-176, Cambridge, MA:MIT Press, 2005.

F. Bissmarck, H. Nakahara, K. Doya, and O. Hikosaka. Efficient learning of real-time motor skills by parallel policies. *IEICE Technical Report*, 104, 23-28, 2004.

F. Bissmarck, H. Nakahara, K. Doya, and O. Hikosaka. Parallel network mechanisms for motor sequence acquisition in real time. *IEICE Technical Report*, 102, 113-118, 2003.

List of abbreviations

BG	basal ganglia
BG-TC	basal ganglia-thalamocortex
CB	cerebellum
CMA	cingulate motor area
CMF	cortical magnification factor
DA	dopamine
DP	dynamic programming
FEF	frontal eye field
GABA	gamma-aminobutyric acid
GP	globus pallidus
GPe	globus pallidus external segment
GPi	globus pallidus internal segment
Glu	glutamate
LGN	lateral geniculate nucleus
LIP	lateral intraparietal area
M1	primary motor cortex

MC	Monte Carlo
MST	medial superior temporal area
MT	middle temporal area
PMd	dorsal premotor cortex
PPTN	pedunculopontine nucleus
RL	reinforcement learning
SMA	supplementary motor area
SN	substantia nigra
SNc	substantia nigra pars compacta
SNr	substantia nigra pars reticulata
STN	subthalamic nucleus
Str	striatum
TAN	tonically active neuron
TD	temporal difference
TMS	trans-cranial magnetic stimulation
Th	thalamus
V1-4	visual cortex areas 1-4
VTA	ventral tegmental area

List of figures

Chapter 1

Introduction

The human motor system that we all take for granted is remarkable in its generality and robustness for the wide range of tasks we perform every day. It is general in the sense that we can reuse experience from a slightly different task; e.g. we have use for experience in squash when we learn to play badminton. It is robust - we can for example grasp and manipulate objects successfully of different sizes, weights and shapes. Furthermore, we have the ability to learn and improve on a wide range of specialized skills over long time.

Despite these abilities, humans and animals face a more difficult control problem than most machines do. Muscles cannot develop force as quickly as electrical motors. Further, biological sensors and relay neurons are often noisy. More importantly, in contrast to the nanosecond time scale of microprocessors, sensory feedback is delayed up to hundreds of milliseconds because of the nature of axonal and synaptic transmission. Because real-time motor behaviour like reaching occurs over fractions of a second, the outcome cannot be confirmed instantly - the brain has to rely on predictions, i.e. assumptions based on prior experience.

Despite the long feedback delays, sensory feedback loops to the central nervous system are ubiquitous and of several modalities (vision, audition, somatosensation, etc), sending information about the state of the external world, and the state of the 600+ skeletal muscles in the human body. These feedbacks must be efficiently used to deliver motor commands in the correct order, at the right time and with the right amplitude. Several movements often have to be coordinated, e.g. for bimanual tasks, or between eye and hand.

In this thesis, we investigate the real-time nature of learning and coordinating multi-modal feedback for complex motor behaviours. Recently, a body of theoretical and experimental evidence has been established, showing that the basal ganglia play a key role of reward prediction, with the potential of optimizing behaviour in this regard. Indeed, the basal ganglia are also multimodal and known to be vital for motor behaviour, including motor procedures. But yet it is unclear to what extent the basal ganglia can optimize real-time motor performance, since latencies are long for the projecting cortical feedback loops. We study the potential of reinforcement (reward) learning of delayed feedback motor control. Further, we study gaze behaviour in a motor sequence learning experiment, to get a better understanding of how visual feedback changes over long time for motor skills, and how feedback is constrained by the physiology of the eye.

1.1. A word on methodology

System's computational neuroscience is a field that tries to formulate sound theoretical bases for learning and control mechanisms of the brain. Theories help to organize and efficiently represent the large body of experimental results. The computational neuroscientist may take two approaches. In the *top-down approach*, she tries to find evidence for a theoretically plausible brain function from available data. In the *bottom-up approach*, she tries to make sense of an experimental finding by a plausible theoretical argument. Both strategies are important means to understand the brain and its disorders. We shall pursue both approaches in different studies in this work (see below).

1.2. Motivation, purpose and scope

The broader aim of this thesis is to understand the long term learning of motor procedures. This class of motor learning often occurs in human ecology; learning to operate a washing machine, typing on a keyboard, or playing the violin. Such tasks are often associated with a speedup of performance [5, 64], which make them a particularly convenient setting to study real-time issues. The difference in performance speed between the novice and the expert may imply a different recruitment of sensory feedback (besides the expert's ability to exploit prediction for control, see below). The sequential nature of motor procedures is also convenient, since the uniqueness of each sequence makes it possible to study the learning process over and over in different examples.

We explore two different topics in this thesis:

I) **A computational study**: In this first part, we extend on a conceptual [62, 63] and computational [123] model, based on experimental work [64] of the basal ganglia-thalamocortical (BG-TC) system [2] and its role in visuomotor sequence learning. In this body of work, it has been established that the prefrontal BG-TC loop and the motor BG-TC loops play differential roles in a cooperative process to achieve robust motor sequence acquisition. Here, we seek to clarify the real-time implications of this system: how feedback delay constrains the performance, and how modalities may be combined. We hypothesize that the BG implement an actor-critic architecture [10, 67, 37]. Our study is further motivated by the fact that in contrast to most other models, we evaluate a 1) real-time and 2) multi-modal learning and control system, which have, as will be shown here, important implications.

   For reasons of limitation, we do not consider the contribution of model-based feedforward control [175] to motor behaviour. Rather, the systems we are considering should be seen as a compliment to feedforward controllers. Consequently, we do not consider the function of the cerebellum, a much important but well investigated real-time control system. Neither should the work be considered an attack on state estimation [163, 97] models; our work is primarily motivated to evaluate the BG-TC system as a real-time motor controller.

II) **An experimental study**: In this second part, we investigate the long-term change in gaze behaviour in sequential hand movements, a topic which has not been explored previously (with the exception of a minor study on macaques [118]). In general, there is a lack of theories of eye-hand coordination [27] . The change of gaze may give important clues on how visual feedback is recruited in later stages of learning, if at all. Also, it has potential to elucidate the utility of peripheral versus foveal vision in motor control, and how gaze-dependent, visual feedback affects accuracy of hand movements.

The subject of eye-hand coordination involves complex computation mechanisms relating the postural geometry of eye, head and arm. We do not consider these computations here - we focus on the spatiotemporal trajectories of gaze relative to manual targets, eccentricity of targets, and the effects of faster execution.

1.3. Outline of the thesis

This thesis is organized as follows: first, important concepts and related fields of research are reviewed. Chapter 2 explains fundamental aspects of theoretical motor control. Proposed control architectures for human motor behaviour are reviewed. Reinforcement learning, an important learning class for achieving goal-driven, optimal motor behaviour is explained. Chapter 3 reviews important experimental findings relevant to the studies of this thesis. The anatomy and function of the basal ganglia are explained in terms of evidence from physiology, focusing on the actor-critic hypothesis. Further, the field of motor sequence learning is reviewed. Here, we particularly look at a series of experiments by Okihide Hikosaka and colleagues, investigating the role of the basal ganglia-thalamocortical system in learning of motor procedures. In the last section, we review previous work in eye hand coordination and gaze strategies of procedural motor tasks. Here we also include early work on the acuity of the eye, and neural mechanisms for spatial representation.

Secondly, in Chapter 4 we present our computational study of real-time, multimodal integration. Related work on feedback control models and modular combination is reviewed. We then propose our alternative model. We demonstrate with two simulation experiments, and report on the results.

Thirdly, in Chapter 5 we present our behavioural study of gaze in sequential hand movements. We review related work on gaze-dependent vision models, representation of neural space, and eye-hand coordination. We present our task paradigm and analysis, and propose a Bayesian model of gaze integration of visual feedback. We report on the learning performance and gaze behaviour of subjects.

Finally, in Chapter 6 we point to potential future directions of our work, and provide a summary of the contribution of this thesis.

1.4. Mathematical notation

Standard linear algebra notation is used throughout the thesis, i.e. vectors are in bold type lower case letters, and matrices in bold type upper case letters. Tilde is regularly used to note an estimated variable of a true variable noted by the same symbol. Superscript T denotes transpose.

Chapter 2

Theories of optimal motor control

One of the most important principles of theoretical motor control is the optimality constraint. Optimization in biology is dictated by survival of the fittest, and works at different time scales by processes of evolution, development, learning and adaptation. Optimization of motor ability is decisive to an organism's fitness; it has direct consequences for foraging, hunting for prey or avoidance of being predated, or searching for a mate.

Accordingly, we can expect that evolution has resulted in 1) control architectures that are optimal for survival and real-time behaviour, and 2) learning mechanisms, i.e. on-line optimization algorithms, effective for constantly readapting to a changing environment. In this chapter, we explain the two important classes of motor control theories, for which almost all biological architectures proposed fit in: closed loop systems, which emphasize optimization of feedback [163], and open loop systems, which emphasize optimization of desired trajectories [175, 88]. We discuss their pros and cons. Then we explain reinforcement learning (RL). While other learning algorithms like unsupervised learning and supervised learning are also important for motor behaviour, it is beyond the scope of this thesis. RL is the learning framework which address the questions how to associate states and actions with respect to reward and task goals for maximal exploitation. Having roots in psychology, it has developed into a formal computational theory with algorithms convenient for on-line optimization. We explain the fundamentals, and focus on the actor-critic algorithm, which is most relevant to biology.

2.1. Control systems

Control theory is a mathematically rigorous field linked to dynamical system's theory, that describes the quality of any control system, in terms of stability, robustness, optimality, observability and detectability. Human motor control is understood by control theory, given the constraints of the brain (the controller), the body (the motor plant) and the environment. There are many constraints much different to those that apply to artificial controllers. For example, sensory feedback loops of the brain are delayed up to several hundreds of milliseconds. The neurons that make up the sensors, inter-neurons and actuators of these loops are noisy and stochastic. Muscles work with a very different dynamic than artificial motors. The environment that humans are exposed to require performance to be robust in a very general sense: for example, objects of different size and shape are manipulated. All these characteristics limit the nature of plausible control mechanisms. By measuring behaviour and brain signals of humans and other primates, a more delicate understanding of human motor control can be acquired.

A very old question is whether humans use feedback or closed loop feedforward control. This question is still relevant [78]. Most neuroscientists would agree that both are important for humans; on one extreme, feedback is certainly used for example when pursuing a slowly, randomly moving object with one's eyes, on the other extreme, we have reflexes responding much faster than feedback can be conveyed. However, as we shall discuss below, the relative weight of either would suggest different control architectures. Theorists dispute whether control is subject to closed loop optimization (corresponding to feedback control) or open loop optimization (corresponding to feedforward control). We shall refer to these as "closed loop control" and "open loop control", although it is not open loop in the strict sense, see below.

In this section, we first introduce the fundamental control problem, and the role of models for prediction. Then, we explain the theory of closed loop and open loop control, and discuss their strengths and weaknesses in terms of theoretical and empirical evidence.

2.1.1 The generic control problem

The fundamental problem we are considering is shown in figure 2.1. A biological *agent* is interacting with its *environment*, characterized by some state vector $\mathbf{x}(t)$. The environment may include both extrinsic (the ground, trees, other biological agents, walls, etc) and intrinsic (states of muscles, motivational states) components. In order to move effectively, the agent needs a control law to select the proper motor command $\mathbf{u}(t)$ for any motor task at any given time t. The sensory feedback signal $\mathbf{y}(t-\tau)$, delayed by τ, is the window through which the agent gets information about the world. The distinction between state $\mathbf{x}(t)$ and feedback $\mathbf{y}(t)$ is motivated, since the agent most often cannot observe the complete state of the environment. Motor control and learning is the science of understanding what internal mechanisms the agent possess to effectively and robustly control and improve performance in this system.

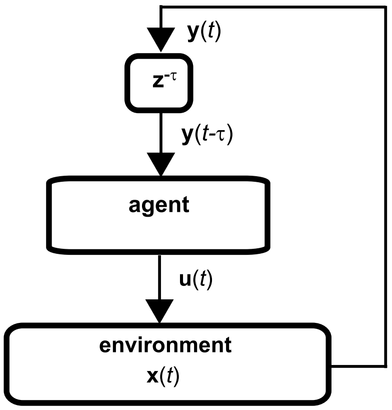

Figure 2.1. The generic control problem.

2.1.2 Models

Prediction is an important concept of control. Predictions based on prior experience can be used for three purposes: 1) to compensate for noisy (unreliable) feedback, 2) to speed up motion when feedback is slow, and 3) to generalize between different motor tasks. Overall, one may say that by letting the brain making predictions (assumptions), information processing becomes probabilistic, and hence more efficient.

For predictions, the brain may utilize internal models, which anticipate the outcome of motor tasks given sequences of responses. Models are used for both open and closed loop control as defined below. There are two classes of internal models: *inverse models* and *forward models*. An inverse model maps the motor command $\mathbf{u}$ given a sensory reading $\mathbf{y}$:

$$\mathbf{u} = F(\mathbf{y}). \tag{2.1}$$

The advantage with inverse models are that they can be applied directly; for a specified desired output (sensory reading) $\mathbf{y}^d$ the motor command is given. In practice, acquiring an inverse model is not trivial, since the mapping is rarely unique (sensory feedback space has higher degrees of freedom than motor command space).

A forward model predicts a sensory reading $\mathbf{y}$ from the motor command $\mathbf{u}$, given the state (mediated by $\mathbf{y}$) of the system:

$$\mathbf{y} = G(\mathbf{y}, \mathbf{u}) \tag{2.2}$$

The forward relation implies a task-specific mapping based on prior experience, where each value of $\mathbf{y}$ typically relates to a unique $\mathbf{u}$. An accurate forward model is memory efficient, and since the next state can be predicted, control is possible without waiting for external feedback, allowing for faster execution.

2.1.3 Closed loop control

In the optimal feedback control paradigm, closed loop control is emphasized. To deal with the long feedback delays and sensor noise, the controller is preceded by a *state estimator* (or *observer*, see Figure 2.2). The role of the state estimator is to predict the state of the environment $\mathbf{x}(t)$ by an estimate $\tilde{\mathbf{x}}(t)$, which is obtained by optimally combining model prediction and sensor feedback by Bayesian inference. The estimate $\tilde{\mathbf{x}}(t)$ is used by a system model (which also may take the motor command as input) to predict the next state, and an observation model to predict the next sensor reading $\mathbf{y}(t)$. The discrepancy between the predicted and observed sensor reading is used to correct the state estimate. For an optimal estimation the gain K should be adjusted so that it increases when uncertainty of the system model is high and the sensor noise is low, and vice versa. For practical implementations, linearity is usually assumed [162, 34], or the computation becomes too expensive in real-time. The model then becomes a Kalman filter [85, 107].

Körding et al. [96] demonstrated that Bayesian inference predicts human behaviour in a partially occluded pointing task, where subjects relied on prior knowledge more when position of hand uncertainty was high. Otherwise evidence for optimal feedback control comes from predicting the *minimum intervention principle*, which states that movements should only be constrained by optimization in task-related directions. Thus it should be allowed to be variable in redundant dimensions. Such variability was demonstrated for some behaviours including hitting, throwing and hand manipulation [163], and obstacle avoidance [103].

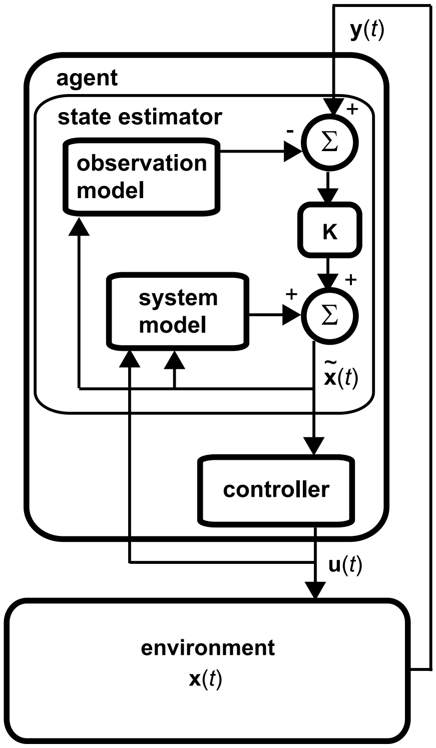

Figure 2.2. A feedback control architecture preceded by state estimation (see text).

2.1.4 Open loop control

Open loop control architectures rely on a planned, desired trajectory $\mathbf{y}^d(t)$ for execution (See Figure 2.3). The desired trajectory is used to compute a motor command $\mathbf{u}^{ff}(t)$ in a feedforward manner. Sensory feedback $\mathbf{y}(t)$ is used as a reference to $\mathbf{y}^d(t)$, to correct any deviation from the desired trajectory by an additional motor command $\mathbf{u}^{fb}(t)$.

In early studies of arm movements, measured trajectories were found to be predicted by explicit cost functions based on arm kinematics, like minimum jerk [49] or minimum torque change [168], which supports the idea of a desired trajectory. Physiological evidence for an open loop architecture [175, 88] much similar to that shown in Figure 2.3 has been identified in neural networks controlling the vestibulo-ocular response (VOR) and the ocular following response (OFR) [53, 95]. These responses are used to keep gaze stabilized with respect to an external scene (VOR) or moving object (OFR), despite body or head rotation.

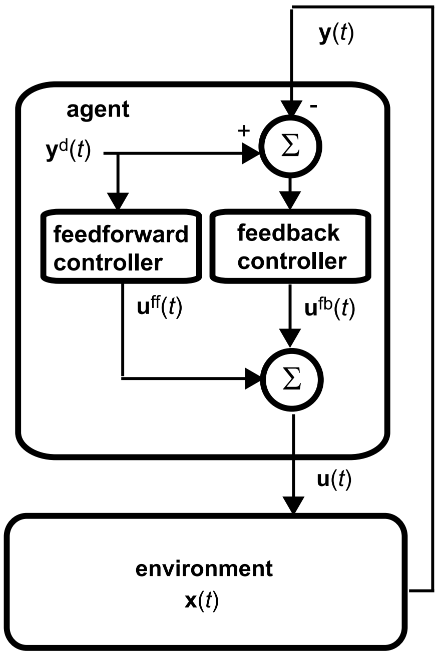

Figure 2.3. An open loop control architecture.

In this case, the desired outcome is explicit - eye rotation should remain invariant relative to the fixated object, so eyes must be rotated counter to the head or body. Cell recording from the cerebellum showed that neural firing of *Purkinje cells* could be well predicted by a linear combination of eye movement kinematics, i.e. the Purkinje cells represented the output of an inverse kinematics model.

2.1.5 The dispute between open loop and closed loop optimization

The optimal feedback control paradigm has recently challenged desired trajectory based control hypotheses [161]. The major advantage of the optimal feedback control hypothesis is that it makes no assumptions that constrain the motor behaviour, other than optimality with respect to task performance. The evidence for that, as discussed above, is motor behaviour is more variable in task-irrelevant directions. Reasons for trajectory based cost functions are harder to find a natural explanation for.

On the other hand, it is a big challenge to explain how efficient learning can be possible under optimal feedback control, given the high dimensionality of real time, dynamic behaviour. Linear control systems are not likely to work well for control in unstable force fields [21]. The long delays of sensory feedback amplify the problem of robust (stable) control of non-linear systems behaviour of walking, for example.

2.2. Online optimization by reinforcement learning

Reinforcement learning (RL) is a theory that comes from experimental psychology, which have origins in the famous classical conditioning experiments by Pavlov [132]. In *classical conditioning*, animals were proven to associate contingent stimuli - preceding states - with reward (or punishment). Reinforcement learning was extended with *operant conditioning* by Thorndike [160]. In experiments of operant conditioning (or *instrumental conditioning*), the animal subject can manipulate the casual chain of stimuli events - states - by different responses - actions. Indeed, animals can also learn to associate sequences of states and actions with expectation of reward. These ideas were formulated in more detailed by Skinner [44]. Later, a mathematical formalism of RL has been developed [94, 16, 153], importing concepts from optimization theory and computer science. From this perspective, reinforcement learning can be viewed as a general, on-line optimization method for complex problems. In the next chapter, we will also see that there are evidence for algorithms of reinforcement learning implemented in the basal ganglia, which is a vital cluster of neural nuclei of the vertebrate brain, also implicated in motor learning and control.

In this section, we review the theory of reinforcement learning. The most important concepts are explained, with an emphasis on the actor-critic algorithm, relevant for neuroscience. Here, we describe a continuous formulation of reinforcement in space and time, which is also used for the framework in Chapter 4. See [38] for a more thorough treatment. For a broader perspective on RL, in the more commonly applied, discrete form, I recommend Reinforcement Learning - An Introduction [153].

2.2.1 The reinforcement learning problem

Consider the system outlined in Figure 2.4. It is no different than the fundamental control problem described in figure 2.1, except that the agent receives reinforcement (reward) $r = r(\mathbf{x}(t))$; a scalar, explicit feedback on what was good or bad for the agent. The function $r(\mathbf{x}(t))$ can have any form, but often have a discontinuous nature in both time and space. In classical RL

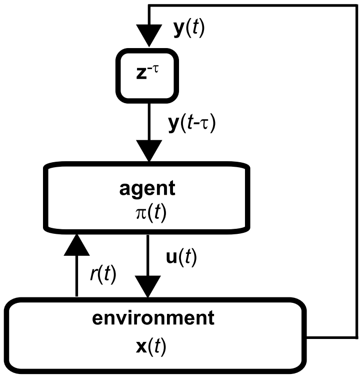

Figure 2.4. The reinforcement learning problem.

problems, $r(\mathbf{x}(t))$ is 0 in every state until some goal state is achieved, e.g. winning a backgammon game [159], or getting to the exit of a maze [153]. In motor control problems [38, 120, 134, 105], $r(\mathbf{x}(t))$ is often augmented by cost terms (negative reward - punishment), or continuous feedback on success. For example, in the former case the square of the joint torque amplitude may be a cost, or in the latter, the agent may be continuously rewarded for not falling [120].

The *policy* $\pi(\mathbf{x}(t))$ is the function of which the agent selects actions $\mathbf{u}(t)$ given $\mathbf{y}(t-\tau)$ and possibly also some internal state variable, like the output of a model. We seek the policy $\pi(\mathbf{x}(t))$ that maximizes the integral of expected future reward $r(t)$. The expected future reward is refered to as the *value function* $V^\pi = V^\pi(\mathbf{x}(t))$, where π indicates the policy the value function corresponds to. The value function is defined as

$$V^\pi(\mathbf{x}(t)) = E\Big[\int_0^\infty e^{-\frac{s}{\tau^{TD}}} r(t+s)ds\Big] \tag{2.3}$$

where τ^{TD} specifies how far into the future returns should be considered. In optimization theory, a Bellman equation [12] states the necessary condition for optimality. In the continuous formulation, the Hamilton-Jacobi-Bellman (HJB) equation [12], a form of Bellman equation, is convenient for theoretical reasons [15]. At time t, the optimal value function V^* is

$$\frac{1}{\tau^{TD}} V^*(\mathbf{x}(t)) = \max_{u(t)\in U}\Big[r(\mathbf{x}(t),\mathbf{u}(t)) + \frac{\delta V^*(\mathbf{x})}{\delta \mathbf{x}} f(\mathbf{x}(t),\mathbf{u}(t))\Big] \tag{2.4}$$

The optimal policy π^* is given by the action that maximizes the the right-hand side of the equation:

$$\mathbf{u}(t) = \pi^*(\mathbf{x}(t)) = \arg\max_{u(t)\in U}\Big[r(\mathbf{x}(t),\mathbf{u}) + \frac{\delta V^*(\mathbf{x})}{\delta \mathbf{x}} f(\mathbf{x}(t),\mathbf{u})\Big] \tag{2.5}$$

The traditional method to find the optimal policy is by *dynamic programming* (DP). In DP, these equations are solved iteratively. First, the value function is evaluated for the current policy (*policy evaluation*). Then, the policy is changed by making π greedier (*policy improvement*), for which a new policy evaluation must be made. These two steps are repeated until convergence.

DP algorithms are important for theoretical reasons, but become unpractical for real world problems. For DP to work, the environment must be perfectly modeled as a Markov decision process (i.e. all state transition probabilities must only depend on the current state, and chance). Another important classical method is the *Monte Carlo method* (MC). In MC methods, the value function is estimated by accumulated experience from simulated or online experience. A model of the environment's dynamics is not needed. A crucial factor is at what rate the policy should be made greedier, a trade-off of exploration and exploitation. A setback with MC methods is that convergence is generally slow, and hard to verify. Experience updating is episodic, as the return is known in the end of the episode. In contrast, many modern RL methods are suitable for learning on-line.

2.2.2 Temporal difference (TD) for reward prediction

Central to RL methods is that the estimation of the value function is updated by the *temporal difference (TD) error*. For the perfect estimation V^{π}, equation 2.3 holds. Differentiating with respect to t, we get

$$\dot{V}^{\pi}(\mathbf{x}(t)) = \frac{1}{\tau^{TD}} V^{\pi}(\mathbf{x}(t)) - r(t). \tag{2.6}$$

As long as the estimation is not perfect, the equality does not hold. The error

$$\delta^{TD}(t) = r(t) - \frac{1}{\tau^{TD}} V(t) + \dot{V}(t). \tag{2.7}$$

is called the TD error. Figuratively speaking, it is the positive or negative surprise at any given time and situation. For prediction, the value function estimation is corrected in the direction of the TD error until convergence to 0.

2.2.3 The actor-critic for control

To improve the policy with TD learning, a number of algorithms have been developed and well investigated, e.g. SARSA and Q-learning [153]. In many of these, action selection is inferred from the value function directly. In actor-critic algorithms, the policy is represented explicitly. Such a differentiation can be efficient when action space is high-dimensional. Also, it seems like a feasible biological implementation (see Section 3.1.3).

A simple actor-critic implementation is outlined in Figure 2.5. The role of the critic is to compute the TD error, which is used to 1) improve the value function estimation $V(t) = V(\mathbf{y}(t-\tau))$, and 2) criticize the preceding action of the actor. The role of the actor is to maintain the policy $\pi(t)$ and adjust actions in the direction of decreasing TD errors.

As an example, assume the policy π is computed by a softmax function

$$\pi_j(t) = \frac{\exp\Big(\beta(a_j(t) + n_j(t))\Big)}{\sum_{j=1}^{J}\exp\Big(\beta(a_j(t) + n_j(t))\Big)} \tag{2.8}$$

where $\pi_j(t)$ is the probability to take action $j = 1, 2, .., J$. The softmax serves to normalize between actions, and to regulate the competition between actions, depending on the inverse temperature β. The term $a_j(t)$ corresponds to the estimated greedy action, while $n_j(t)$ is an exploration term. For efficient learning, we want to reinforce, or penalize the action deviation $E_j(t)$ from the greedy action:

$$E_j(t) = \frac{(\pi_j(t) - \hat{\pi}_j(t))^2}{2} \tag{2.9}$$

where the circumflex in $\hat{\pi}_j(t)$ denotes the greedy action (i.e. with noise term $n_j = 0$).

2.2.4 Eligibility traces for efficient actor-critic learning using function approximators

A problem in reinforcement learning is that of credit assignment - how to know which actions caused the observed reward. If no bias can be applied beforehand, which is typically the case, the proper actions will be found statistically by

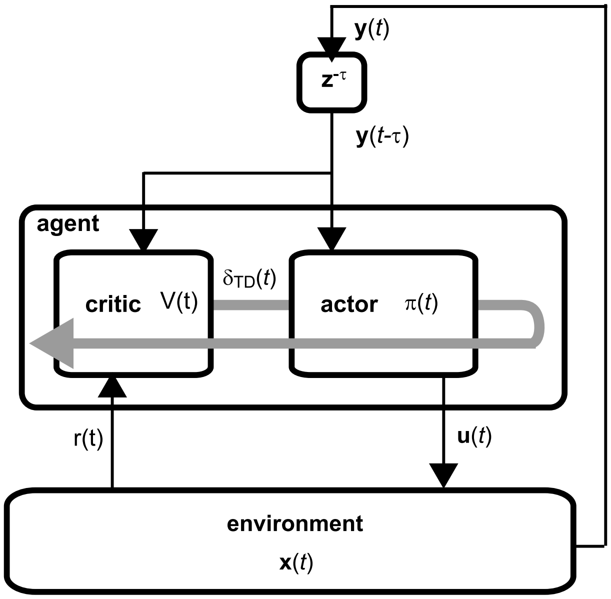

Figure 2.5. The actor-critic architecture. The critic computes the TD error $\delta_{TD}(t)$, the discrepancy between actual and expected reward. This signal is used both to improve reward prediction of the critic, and to criticize the preceding action $\mathbf{u}(t)$ by updating the policy $\pi(t)$ (see text).

experiencing many trials. *Eligibility traces* are memory traces of applied actions and states that help temporal credit assignment. While eligibility traces may have an arbitrary form, they typically decay exponentially from the time the state was visited or the action was taken. Assume that we are using function approximators $V = V(\mathbf{w}^c)$ for the critic, and $\pi = \pi(\mathbf{w}^a)$ for the actor, as we shall in Chapter 4. Function approximators are typically trained by stochastic gradient descent [154].Then, the traces reflect the recent history of function gradients with respect to parameters $\mathbf{w}$, rather than states and actions explicitly. For the critic, the exponential eligibility trace e_k^c for parameter w_k becomes

$$\dot{e}_k^c(t) = -\frac{1}{\tau^{ET}} e_k^c + \frac{\partial V}{\partial w_k^c} \tag{2.10}$$

and for the actor eligibility trace e_k^a

$$\dot{e}_k^a(t) = -\frac{1}{\tau^{ET}} e_k^a + \frac{\partial E_j(t)}{\partial w_k^a} \tag{2.11}$$

where parameters w_k^a reflecting the action deviation are stored for reinforcement, and the time constant τ^{ET} determines how far back in time states should be included. The trace for the actor is given from

$$\frac{\partial E_j(t)}{\partial w_{kj}^a} = (\pi_j(t) - \hat{\pi}_j(t)) \frac{\partial \pi_j(t)}{\partial w_{kj}^a}. \tag{2.12}$$

The update learning equations for the parameters with gradient descent become

$$\dot{w}_k^c = \alpha \delta^{TD}(t) e_k^c(t) \qquad \dot{w}_{kj}^a = \alpha \delta^{TD}(t) e_{kj}^a(t) \tag{2.13}$$

where α denotes the learning rate.

2.2.5 Policy gradient methods - an alternative

Actor-critic algorithms may run into problems in complex learning tasks. If the value function is represented by a function approximation, convergence cannot be guaranteed. Also, if a large part of the state is hidden, actor-critic methods are known not to behave well. Recently, a different class of reinforcement learning has regained interest, called policy gradient algorithms [174, 154, 81]. Here, the policy $\pi(\mathbf{w})$ is updated directly by estimating the policy gradient with respect to the parameters $\mathbf{w}$, in the direction of increasing average reward. Such policy improvement is guaranteed to converge [154]. However, with conventional gradient descent, convergence is typically very slow for real world applications. Using the natural gradient [3] was shown to make learning more efficient [81, 134, 105, 133]. Learning is typically episodic like MC, since the average return must be computed. The high variance between samples often observed is a problem with policy gradient methods. Peters et al. proposed “the natural actor-critic” [134, 133], where the critic is used to for computing the natural gradient, and reducing variance.

2.3. Conclusion

For optimal and robust motor behaviour, the brain needs the benefits of two extremes. On one hand, there are fast, reflex-like, feedforward driven control mechanisms which can deliver the motor commands well in real time, but with little capacity to learn and utilize from extrinsic modalities like vision or audition. On the other hand, there are high level, plastic, multi-modal feedback loops which can learn to optimize goal-driven behaviour, but are challenged by their long feedback delays for effective motor control. While both open loop and closed loop control architectures have been successful for explaining particular behaviours, it remains to be seen whether both mechanisms are present in the brain, or if an integrated control mechanism exists.

Further, for optimization of motor behaviour the RL framework is a promising framework that may successfully explain biological learning. However, for RL to be feasible, satisfactory solutions to at least two problems have to be explained. First, learning algorithms must be efficient enough to overcome the inherent slowness of trial-and-error learning. This includes the problem to set the appropriate meta parameters; learning rate, exploration-exploitation trade-off, and temporal discounting of reward. Second, the reward signaling systems of the basal ganglia are far from the motor plant, and can only influence behaviour by feedback loops with considerable delay. These problems are yet poorly understood and remain a challenge for theorists.

Chapter 3

Neural learning and coordination of visuomotor procedures: experimental evidence

In the previous chapter, we looked at general theories for human motor control and learning. In this chapter, we review experimental work of psychology and neurophysiology about multi-modal learning and control of motor behaviour. We focus on psychological and physiological work particularly important for the studies of this thesis. First, we review the anatomy and function of the basal ganglia, including evidence for reward prediction mechanisms. Then, we look at the physiology of goal-driven motor sequence learning, where we particularly review the work of Hikosaka and colleagues [62, 65], which is important for understanding the motivation of the study in Chapter 4. In the following section we look at eye-hand coordination and gaze strategies. We first explain physiological factors that we believe influence gaze strategies, i.e. neural mechanisms of spatial representations and properties of acuity of the retina. Then we review the little work that has been done on gaze behaviour for procedural movements and skill learning, which is important background for Chapter 5.

3.1. Goal-driven learning of motor sequences

Many brain areas contribute to procedural motor learning. Here, we focus on the role of the basal ganglia-thalamocortical system. The basal ganglia are particularly interesting, because they seem to perform some integrating function from vast cortical input, and also signal reward information. We first review basic anatomy of the basal ganglia loop circuit, and then experimental findings and models of the role of BG in reward prediction and action selection. We then review work on motor sequence learning in BG-TC loops, with a more thorough explanation of the parallel loop hypothesis [62] and its underlying body of work.

3.1.1 Basal ganglia: a paradigm shift

The basal ganglia are a subcortical structure of deep nuclei, linking the cerebral cortex to the thalamus. It is a phylogenetically old system, relatively well conserved among vertebrates, both structurally and pharmacologically. Traditionally, the function of the basal ganglia has been attributed to motor control, as its associated neural disorders like Parkinson's disease causes motor deficits. Recently, this view has been challenged, as there is evidence for implications of the basal ganglia in mood, cognition, and non-motor behaviour [86, 37]. From a computational viewpoint, the actor-critic model has been advocated by many [10, 119, 36, 31] to capture the function of the basal ganglia-thalamocortical (BG-TC) system as a framework for reinforcement learning and action selection.

In Chapter 4, we shall evaluate a real-time actor-critic framework for multimodal control and learning of motor skills. To put the actor-critic model in a neurobiological perspective, we review the neuroscience of the basal ganglia in this chapter.

3.1.2 Basal ganglia: anatomy

The basal ganglia are usually defined to include the following nuclei: *the striatum* (Str, including the *putamen*, *caudate nucleus* and *nucleus accumbens*), the *globus pallidus* (GP), the *subthalamic nucleus* (STN), and the *substantia nigra* (SN) (see Figure 3.1). Sometimes the *ventral tegmental area* (VTA) is included.

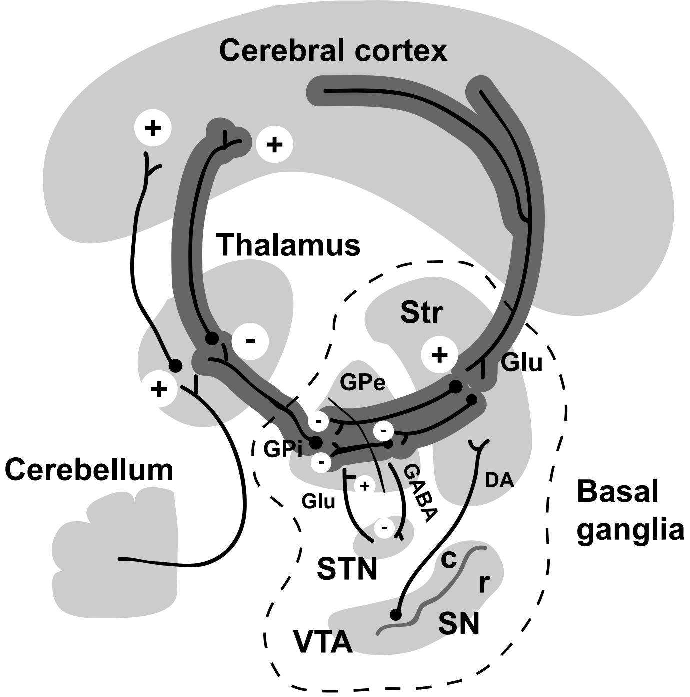

Figure 3.1. Anatomy of the Basal ganglia-thalamocortical loop. The main loop (highlighted in dark gray) consists of cortical glutamate (Glu, excitatory) projections to the Striatum (Str), inhibitory convergent Str neurons to the Globus Pallidus (GP), inhibitory GP neurons to the thalamus, which relays back (excitatory) to the cortex, and also receives input from the cerebellum. The loop is branched into direct and indirect pathways through the GP (see text); the direct goes through the internal segment (GPi) only, while the indirect is relayed through the external segment (GPe). Dopaminergic (DA) neurons project to the Str from the substantia nigra (SN) pars compacta (c) (r - reticulata, the other compartment of SN). Reciprocal connections exist between the GP and the subthalamic nucleus (STN, *gamma-aminobutyric acid* (GABA) is an inhibitory neurotransmitter). The ventral tegmental area (VTA) also has efferent dopamine neurons. Adapted from *Graybiel* [57].

The striatum is the input stage of the basal ganglia. The main cell type of the striatum is the *spiny neuron.* It receives input from *glutamatergic neurons* of the cerebral cortex. The spiny neurons are organized into *striosomal* and *matrix* modules. They differ in their chemical makeup, but also in their efferent projections. The matrix neurons target the internal segment of the globus pallidus, while the striosomal neurons target *dopaminergic* (DA) neurons in the SN and VTA. The striatum also has interneurons called tonically active neurons (TANs).

The globus pallidus consists of two segments, the external (GPe) and internal (GPi) segments. Both receive input from the striatum. The GPe targets GPi, which in turn target the thalamus. There are thus two possible pathways through GP: the *direct pathway*, through GPi only, and the *indirect pathway*, through the GPi via GPe. The substantia nigra has two compartments: *substantia nigra pars reticulata* (SNr) and *substantia nigra pars compacta* (SNc). The SNr is similar to the GPi, though it is closely innervated with the SNc. The afferent input to the SNc is complex, but the major sources are the striatum and the *amygdala.* The SNc sends dopaminergic input to the striatum. The VTA is a dopamine-rich nuclei close to the substantia nigra, targeting the *nucleus accumbens.* The STN has recurrent connections with the GP.

To summarize the functional connectivity, the cerebral cortex, Str, GP and the thalamus form a recurrent loop, which is highly convergent: the ratio of cortical afferent and pallidal efferent neurons of the striatum is about 80:1 in the monkey and 30:1 in the rat [56]. The GP has a high, spontaneous firing rate (50 Hz), which is inhibited by striatal input. The GP in turn inhibits the thalamus excitatory input to the cerebral cortex. The possible advantages of this double negative signaling over a single positive are not well understood.

The SNc, VTA and STN are considered to subserve this BG-TC loop rather than being a part of it. How the entire system may function is discussed below.

Further, the cortical input to the striatum is modular in its topography: The projections have been identified to four distinct BG-TC loops, originating in distinct cortical areas: the limbic, prefrontal, oculomotor and motor loops [2, 114]. These parallel circuits do not seem to converge with one another, but rather project back to their origin [2]. Each loop is also topographic. For example, the motor loop is somatotopic throughout its cortical, striatal, pallidal and thalamical parts. This modularity is suggested to reflect parallel processing, of reward prediction [155] and motor procedures [62] (see Section 3.1.4).

3.1.3 Basal ganglia: function

Until the 1980's, most knowledge about basal ganglia function was from clinical observations of Parkinsonian and Huntingtonian patients. The obvious behavioural change caused from these neurodegenerative disorders was impairment of voluntary movements. The basal ganglia was believed to be directly involved in the production of movement. It was referred to as the "extrapyramidal" in contrast to the "pyramidal" motor system, as it was not directly connected to the spinal cord.

However, recently it has also been recognized that cognition and mood are affected by these disorders [86, 37]. Also, many other areas, such as the premotor cortex, motor cortex and cerebellum contribute to movement, and lesions in these areas consequently also lead to motor impairment.

The actor-critic hypothesis

For the last 15 years, theorists have been trying to find a feasible computational function of the BG-TC system [68, 37, 57, 89]. Although the four BG-TC loops project cortical areas with very different functions, the structure of the basal ganglia is highly conserved across the loops, and is likely to perform the same operation on each loop. This suggests an abstract function plausible for cognitive, emotive, motor and oculomotor circuits throughout.

The most common class of computational models of the basal ganglia is the actor-critic model, a class of reinforcement learning algorithms (see Section 2.2). Recent experimental discoveries have led to this paradigm shift in basal ganglia function. Particularly important observations are, 1) the response of the SNc seems to be related to expected reward, like a temporal difference signal (see below), and 2) the high convergence in the BG loops suggest a state (high dimensional) to action (low dimensional) mapping.

In ground-breaking experiments, Schultz and colleagues [147, 148] recorded the response of SNc neurons of monkeys behaving in a simple, classical condi-

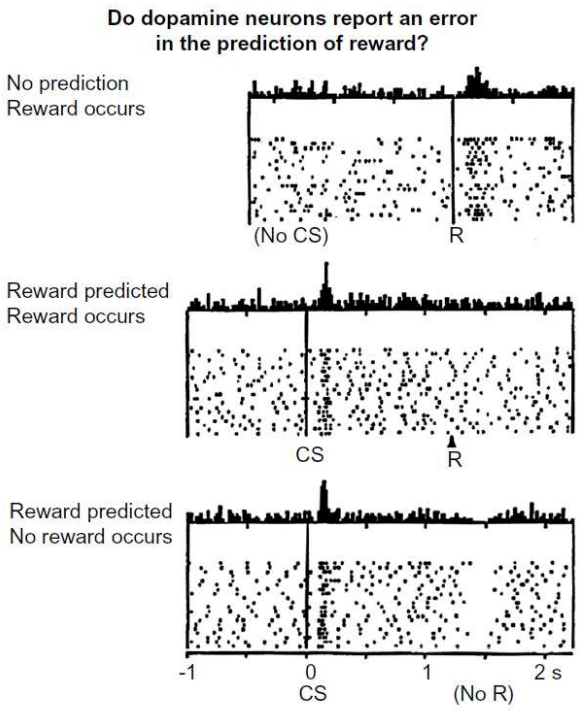

Figure 3.2. Reward prediction by dopaminergic neurons of the substantia nigra. The recorded neuron fires when reward cannot be predicted (top) just after delivery (R). In presence of a contingent stimulus (CS), there is no response to the reward itself, but instead to the associated CS (center). When reward is neglected for the CS associated with reward, the baseline activity is depressed shortly after expected reward delivery (bottom). From *Schultz et al.* [148].

tioning task (see Figure 3.2). For unexpected reward, SNc neurons would respond at the time of reward delivery (a positive surprise). For anticipated reward, the response would come at the preceding cue indicating reward. Also along with the hypothesis, activity was depressed at the time of reward when reward was omitted (a negative surprise).

The signal observed is very much like the temporal difference error in reinforcement learning, signaling the deviation from reward expectation (see Section 2.2). A controversial topic is whether there exists a temporal difference-like signal for negative reward, i.e. punishment. Daw et al. [32] suggested serotonin as a plausible agent for negative temporal difference error.

Wolfram Schultz' discovery of a temporal difference-like neural response paved the way for proposing that the basal ganglia implements the actor-critic. The nominal neural model was introduced by Houk et al. 1995 [67] (see Figure 3.3). They proposed that *striosomal modules* served as critic (striosomes in Str, STN, and DA neurons) and the *matrix modules* as actor. The TD error in the model is computed from three sources corresponding to the three terms of the TD error. Primary reward is assumed as input from the lateral hypothalamus. The instantaneous value of the value function is provided by direct, tonic inhibition of the SNc from striosomes, while a phasic excitatory, delayed input corresponding to the value function at the previous time step is relayed via the STN. The model could account for reward prediction in Schultz' experiment ([147] see above), but not to reward cancellation. The model did not provide any details or arguments for the actor, other than the observation that matrix modules target the output stage of the basal ganglia, the GP neurons.

Later models did account for a more timing-sensitive mechanism of the critic, accounting for reward cancellation ([119, 152, 37]. However, these models make assumptions that are not supported by neurophysiological evidence. Recently, Kawato & Samejima [89] proposed that the pedunculopontine tegmental nucleus, a nucleus (PPTN) in the brainstem, takes part in the computation of the value function. The PPTN is not considered to be part of the basal ganglia, but projects excitatory neurons to the substantia nigra [54], and its lesion causes Parkinsonian-like tremors [110].

Much less work has been done to explain the actor in terms of known neurophysiology [74], although some broadstroke attempts have been made [67, 14, 9, 123, 18]. Little is known how the outputs of the basal ganglia are affecting the cortical network, how outputs are integrated and what they represent [57].

3.1.4 Motor sequence learning systems

In the early 1960's, Paul Fitts proposed that procedural skills are learned in three stages [47, 5]. In the first *cognitive stage* the subject must learn the nature of the task by declarative instructions or examples, and be attentive to sensory

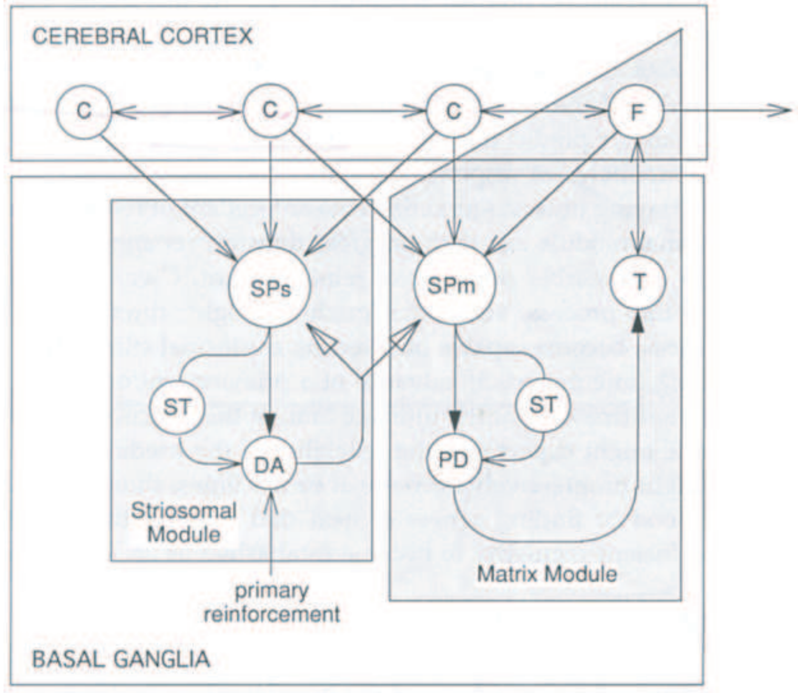

Figure 3.3. The actor-critic model of Houk et al. The striosomal module implements the critic. Cortical neurons (C) project the state to the striatal spiny neurons (SP) needed to compute the value function. The temporal difference error is computed by dopaminergic (DA) neurons by referencing primary reinforcement and the value function at the previous time step, a signal relayed from the striosomal spiny neurons (SPs) through subthalamic nuclei (ST), with the tonic inhibiting signal projecting directly from the striatum. Matrix spiny neurons (SPm) learn by the critic to mediate the action by double inhibition through pallidal (P) and thalamic (T) neurons, for prefrontal (F) neurons. From *Houk et al.* [67].

feedback. In the second *associative stage* she learns to proceduralize the skill, so that movements are recalled by association of the previous. Here, performance becomes more fluid and accurate. In a third *autonomous stage* the subject can perform the skill more and more accurately and rapidly. The cognitive involvement gradually diminishes, and she may attend to other simultaneous tasks [5].

The dichotomy between the nature of the early cognitive stage and the late autonomous stage suggests that there are different neural mechanisms involved in respective phases of learning. The cognitive stage is characterized by slow execution, attention, declarative knowledge, extrinsic feedback and closed loop control. The autonomous stage is in contrast characterized by fast execution, unconscious recall, procedural skill, effector-specific (e.g. right hand) specialization, intrinsic feedback and open-loop control. There must not be only mechanisms supporting learning for different stages, but also for coordination and transfer between systems.

In the 1990's, Okihide Hikosaka designed an experimental paradigm for the purpose of studying the psychology and the physiology of motor sequence learning, called "the 2 x 5 task", which resulted in many novel findings [64] (see Figure 3.4). In this experiment, subjects execute a sequence of reaching movements to press 10 (2 x 5) keys on a panel, where they improve both in terms of accuracy and speed, over a long time (several days of practice). The design of the 2 x 5 task was motivated by 1) the fact that it was simple enough for a monkey to perform, but yet not too simple, and 2) the possibility for single subjects to learn many sequences, as some 10^{10} combinations are possible, and the learning of each sequence is unique. The hierarchical structure of the sequence was also motivated by the reminiscence of real-world procedures.

Behaviourally, subjects of the 2 x 5 task improved on accuracy (errors) across a single day, while improvement in performance time (time to complete the trial) improved across several days [64, 63]. There was also some task learning (improvement of initial performance for every new sequence). Learned sequences were also found to have a retention time of at least 6 months, as monkey subjects performed better for sequences learned 6 months in advance than for new sequences [64, 65].

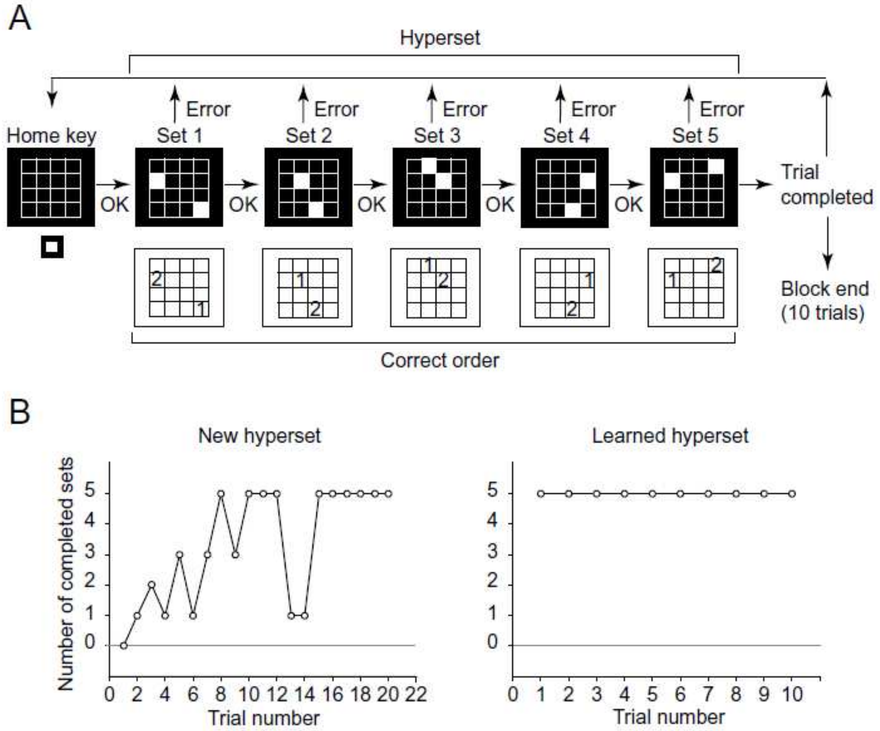

Figure 3.4. The 2 x 5 task. (A) Subjects press a home key below the key panel (4 x 4 grid) to start the trial. 2 keys, called a *set*, is presented (black background) with a hidden order (shown on white background). If the wrong second key is pressed, the subject must start the trial over again, otherwise she may press the second key to proceed to the next set. A trial is completed by five sets, which also completes a hyperset. The exact same hyperset is presented over and over until the subject has completed 10 successful, consecutive trials. For monkeys, reward (juice) is given by each completed set, with increasing amount across the hyperset. The human version of the task is performed by 10 sets in the hyperset. (B) Typical accuracy performance (completed sets) of a new hyperset (left) and a learned hyperset (right). From *Hikosaka et al.* [62].

Medial motor cortex

While experimental studies have shown that many brain areas are involved in procedural learning [55, 66, 87, 164, 121], the medial motor areas, including the *supplementary area* (SMA), seem to play an intricate part [157, 136]. While the primary and premotor cortices respond to single limb movements [84], those of the medial motor areas are sensitive to an embedded motor context [86, 157]. Pioneering work by Jun Tanji first distinguished the *pre-supplementary motor area* (pre-SMA) from the SMA proper [106]. While the two areas are reciprocally connected, the pre-SMA projects to the prefrontal cortex [156, 106] and the rostral cingulate motor area (rCMA) [11]. Tanji's group also demonstrated functional differences between the areas recording monkeys performing overtrained motor sequences. pre-SMA neurons responded specifically to initiations and rank order specific movements, while SMA neurons were context specific, but active during ongoing movements [157]. In this regard, the cingulate motor areas (dorsal and ventral) have also been shown to have response properties much similar to the SMA [141].

In a recent trans-cranial magnetic stimulation (TMS) study, Kennerley et al. showed that in a chunked sequence, lesion of the pre-SMA by TMS caused a longer reaction time, while TMS pulses within the chunk had no effect [91].

In the 2 x 5 task, the pre-SMA was shown to respond preferentially to novel sequences compared to learned sequences [126]. With impairment of the pre-SMA by muscimol injection, learning of novel sequences was impaired, but no disruption of learned sequences was observed.

Cerebellum and basal ganglia

Both the cerebellum and the basal ganglia are involved in motor sequence learning. The cerebellum is important for real-time control of motor sequences. Lesions of the cerebellum impair motor sequence learning [115], but not conditional visuomotor learning [128]. In particular, the cerebellum may be involved in learning and storing internal models of complex motor skills [69], used for feedforward execution. In the 2 x 5 task, lesion of the output stage (*dentate nucleus*) of the

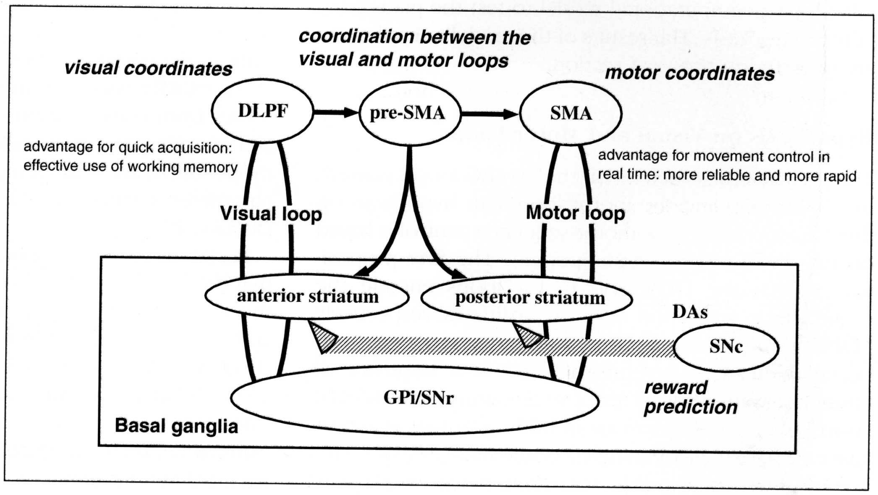
coordination between the visual and motor loops
visual coordinates
motor coordinates
DLPF
pre-SMA
SMA
advantage for quick acquisition: effective use of working memory
advantage for movement control in real time: more reliable and more rapid
Visual loop
Motor loop
anterior striatum
posterior striatum
DAs
SNc
GPi/SNr
reward prediction
Basal ganglia

Figure 3.5. The parallel loop hypothesis. The visual (prefrontal) loop learns the sequence in visual coordinates in the dorsolateral prefrontal (DLPFC) cortex, in parallel with the motor loop, learning the sequence in motor coordinates in the SMA. The basal ganglia implement reward-driven learning and action selection by an actor-critic implementation. The loops are coordinated by the pre-SMA. See text for further details. From *Nakahara et al.* [123].

cerebellum disrupted performance of learned sequences, but not the learning of novel sequences [104].

Many studies have shown implications of the basal ganglia in procedural learning [80, 165, 62, 75, 121]. In the 2 x 5 task, the neurons in the anterior striatum were preferentially responsive to new sequences. In the posterior putamen, neurons responded preferentially to learned sequences [116]. Analogously, inactivation of the striatum affected mostly new sequences in the anterior striatum, while it only affected learned sequences in the putamen [117].

The parallel loop hypothesis

Basal ganglia-thalamocortical loop circuits are organized in a modular fashion, including the prefrontal and motor loop [2, 114] (see Section 3.1.2). Given the observations reviewed above of the pre-SMA and anterior striatum (part of the prefrontal loop) preferentially involved in learning of new sequences, and the posterior putamen (part of the motor loop), preferentially involved in learned sequences, led to the parallel loop hypothesis [62, 65], which states that the prefrontal loop learns new sequences, and the motor loop stores and executes well practiced sequences.

In a model by Nakahara et al. (Figure 3.5), computational mechanisms of the parallel loop hypothesis and their possible advantages were analyzed [123]. It was proposed that the prefrontal loop, now called "the visual loop", represented sequences in lower-dimensional, visual (external world) coordinates, while the motor loop stored sequences in high dimensional motor (e.g. joint angle) coordinates. The two loops contribute differentially to the output: the visual loop by learning speed, enabled by the lower dimensionality, and the motor loop by accuracy, enabled by the higher dimensionality. The basal ganglia learn the sequences by the actor-critic algorithm (see Sections 2.2 & 3.1.3). The pre-SMA was hypothesized to work as a coordinator between the loops, and the (ventral) premotor cortex worked as a translator of motor commands output by the visual loop to realize motor commands in the motor loop.

With the model, Nakahara et al. could show that learning with both loops in parallel was more efficient than any of the loops alone. Simulation results analogous to experimental results [117, 116, 127] of the 2 x 5 task could also be shown. It was shown that blockade of the visual loop affected new hypersets more than learned, while the difference was less for blockade of the motor loop. Blockade of the coordinator (corresponding to pre-SMA) affected learned sequences but not new.

The model implementation by Nakahara et al. was step-wise movement-by-movement, and two dimensional in both visual and motor coordinates. Other interesting problems would arise with a more realistic model. A high degree of freedom arm would give rise to the inverse problem in the visual loop, i.e. each visual coordinate would map to a space of possible motor coordinates. Visual and somatosensory feedback are known to have different latencies, which makes it challenging to coordinate the two loops (see Chapter 4). Another issue that may be more difficult in a high dimensional system is how to integrate the loop outputs. The model successfully evaluated and learned actions locally in the two loops, as the anatomy suggests [114]. That implies a multi-agent system, where the loops must be informed efficiently about the state and outputs of the other loops, or a correct integrated output may not be coordinated. It remains unclear how output of BG-TC loops are integrated [60].

3.2. Eye-hand coordination of motor procedures

An important aspect of understanding gaze strategies for manual skills, is the fact that eye and hand movements are coordinated. This implies more than the intuitive notion that the eye refixates to provide visual feedback for hand movements. First, in reaching and pointing, hand and eye movements are triggered by the same neural systems. This is expressed behaviourally by a fixed latency of 60-100 ms, depending on the task, between initiation of hand and eye movements [138, 46]. For effective coordination, eye and hand share neural representations of space [24, 27]. This implies a complex system of coordinate transformations, from the low-level muscle states and angular coordinates of the arm, to head-centered, body-centered, world-centered and retinal coordinates. The brain seems to maintain these representations in parallel [24] . This requires extensive complex computations, as the relationships must be updated with any rotation of body, head and eye.

In this section, we look at aspects of vision related to eye hand coordination and gaze strategies for manual procedural skills. First, we review what is known about neural mechanisms of spatial representations, which are presumably important for computing pointing movements, for example. Then, the acuity of the eye with respect to eccentricity is explained, a factor that may be important for gaze strategies, as argued in Chapter 5. Finally, we review the few studies that have been made of gaze strategies for manual sequences in controlled experiments.

3.2.1 Neural representations of space

Monkey neurophysiology has provided important insights how neural systems are maintaining spatial representations. In this system, the parietal area is playing a key role in maintaining spatial representations for potential action (an extensive

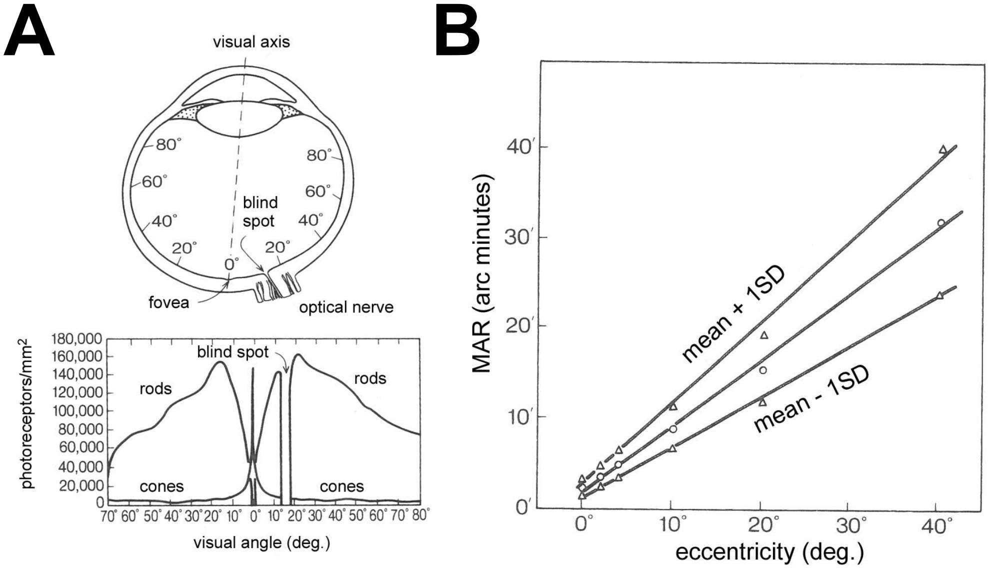

Figure 3.6. Visual acuity depends on retinal eccentricity. (A) Density of cones and rods in the retina. Cones are almost exclusive to the fovea around zero degrees. Rods are sparse in the fovea, but also decrease in concentration in the off-foveal range. (B) The minimum angle resolution (MAR, in arc minutes, corresponding to the Landholt C test) is linearly increasing with retinal eccentricity. From *Weymouth* [173].

review is given by [24]). For eye-hand coordination, the *lateral intraparietal area* (LIP) is of particular interest, as it is believed to represent the space explored by eye movements, in retinal coordinates. The LIP, together with the *frontal eye field* (FEF) [20], the prefrontal cortex [52], the *superior colliculus* [171] and *extrastriate areas* [124, 125] takes part in *remapping* of stimuli [41, 13]. A stimulus flashed before a saccade is remapped after the saccade to bring it to the new retinal location. Such a memory trace of the location of a previous stimuli can last up to minutes [167]. These spatial memory updates may be important to understand gaze behaviour.

3.2.2 Acuity and eccentricity

The need for a gaze strategy arise from the property of the retina: the acuity of the fovea is far better than that of the periphery. This is reflected in the physiology of the eye. The fovea has a higher density of photoreceptors than the periphery [173] (see Figure 3.6A). The density of cones in the fovea is approximately 40 times that of the periphery [29, 28]. The foveal input is further magnified in each projection layer of the early afferent visual pathway: compare to the peripheral projection, there are 4 times more ganglion cells per cone, 4 times more *lateral geniculate nucleus* (LGN) cells per ganglion cell, and 10 times more *striate* cells per LGN cell [25]. In the primary visual cortex (V1), the *cortical magnification factor* (CMF) is defined as the ratio of projected retinal angle to mm of visual cortex [30]. CMF is significant because it has been shown to scale with properties of the eye, most importantly several measures of acuity [26, 35, 102, 150, 43, 42].

Visual acuity is dependent on retinal eccentricity. A linear positive correlation to retinal eccentricity has been found for acuity measures such as minimal angle of resolution (MAR, corresponding to Landholt C, see Figure 3.6B) [173], Vernier, and grating acuity [42]. To demonstrate this property, Anstis showed that letters and images can be made “equally readable” [8] and “equally blurry” [7], respectively, by scaling font size and image blur by radial distance from the center of a fixation point.

The linear dependence of eccentricity is also reflected in accuracy measures of pointing [138, 17, 158]. Rossetti et al. found a quadratic dependence of pointing surface error (end point variance) in an eccentricity range of 0-40 degrees [140]. In pointing, accuracy is also dependent on eye and head rotations (see [27] for review), for which eccentricity-dependent accuracy has been found independent [140].

3.2.3 Gaze strategies of manual skills

In various human behaviours in sports and daily activities, gaze strategies have been shown to be highly specialized for the task at hand, for example in typing [70], driving [99, 100] and batting [101]. Often, it can also be observed that professionals, for example in golf putting [169] or basketball throwing [170] have a more effective, specialized gaze strategy than novice players. However, in general, apart from the fact that targets are fixated pro-actively [138, 46, 76], there is a lack of general theories of gaze strategies and eye-hand coordination for manual tasks.

Recently, however, a series of experimental studies of Johansson and colleagues in controlled environments have led to important insights of gaze strategies for manual tasks [76, 158, 142]. Johansson et al. first studied gaze in a simple, sequential manipulation task [76] (see Figure 3.7). Subjects had to grasp a rectangular bar, avoid an obstacle during an aimed reaching movement, put the bar down at a target site, and then reverse the procedure. It was shown that subjects fixated important landmarks - primarily grasp sites, secondarily obstacles. Never did subjects fixate the hand or moving bar. Also, it was shown that when subjects observed others performing the same task, they displayed a gaze behaviour much similar to that of the active subject, gazing predictively at hand targets [48]. This finding supports the idea that action understanding results from predicting the observed action from one's own motor representation.

In another study, subjects were instructed to point to four targets on a screen with a pen after visual presentation [158]. The visual presentation was either four targets only, targets and distractor targets, or four targets with a sequential instruction (arrows between the targets). Presentation time was variable (0.5-8 s), after which the screen went blank, and subjects could start pointing. The study showed that with targets only, subjects could effectively remember the position of targets in parallel by peripheral vision. The sequential instruction further enhanced performance. With distractors, however, targets had to be foveated, making the performance error dependent on presentation time and number of fixations.

Also a study of gaze behaviour for a visuomotor learning task was done by the same group [142]. Subjects had to learn to control a cursor on a screen rotating a manipulator in two dimensions, requiring a non-trivial visuomotor transformation. Learning occurred in three stages. In a first exploratory stage, the performance was erroneous, and gaze was responding reactively to the cursor. In the second, gaze was directed targets, and the (manual) success of directing the cursor to the target improved. In a third refinement stage, gaze was shifted

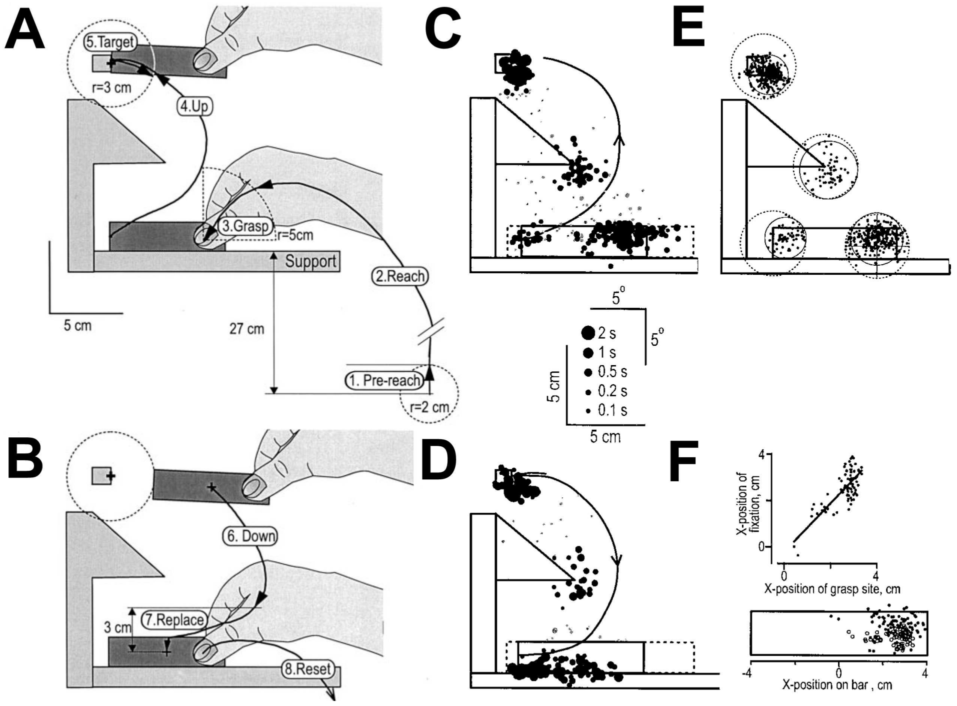

Figure 3.7. A study of gaze strategies for a simple procedural manipulation. (A-B) Experimental procedure. Subjects were asked to grasp the bar, put it on the top surface, avoiding the protruding obstacle (A, "up-phase"), and then reverse the procedure (B, "down-phase"). (C,D) Distributions of gaze fixations in the up-phase (C) and in the down-phase (D). Fixation dots are scaled for duration. (E) Distributions of fixations (solid circles, center indicate mean) with respect to landmark zones (dotted circles, center indicate mean). (F) Correlation between fixation location and grasp location on the bar (top). Below fixations (dots) and grasp sites (circles) are displayed with respect to the bar. Adapted from *Johansson et al.* [76].

directly toward the target, while manual performance gradually improved. The experiment well demonstrated how subjects must learn a predictive mapping between manual actions and eye movements.

Gaze behaviour was also briefly investigated in the 2 x 5 task [118] (see above). It was found that saccades gradually became predictive, fixating the first correct target prior to illumination (presentation) of two targets. the likelihood of such saccade increased over 20-30 days. Use of the opposite hand disrupted the anticipatory saccades.

3.3. Conclusion

The basal ganglia are projecting to prefrontal areas, motor areas, oculomotor areas, and the limbic system. They are also involved in reward prediction. Lesions of the basal ganglia cause motor disorders. Thus, the hypothesis that the basal ganglia are involved in multimodal integration of goal-driven motor behaviour is well justified. However, a lot of experiments have yet to be done before a more clear idea about the basal ganglia function can be obtained. It is known that dopamine neurons signals reward expectancy, but we have no clear evidence for how the value function (the reward prediction) is computed. It is still a mystery how the BG-TC modular loops are coordinated and integrated, and how their outputs are mediated to actual muscle actuation.

From a computational viewpoint, we believe it is time to consider the real-time nature of signals in the BG-TC system. There may be several temporal constraints that determine if the actor-critic hypothesis, for example, is feasible. The long delays of extrinsic (visual, auditory, or somatosensory) feedback impose a challenge for the brain to coordinate modalities and use feedback efficiently for controlling motor behaviour.

In this context, we find it important to understand what the eye and hand do in fine spatiotemporal detail. For example, as we shall see in Chapter 5, vision seems to play a part for guiding also for very well rehearsed hand movements, even though the extrinsic environment is static and certain. How could vision contribute to accurate performance, when control based on predictive components or intrinsic feedback, including proprioception, is well trained? To understand that we must know the precise nature of visual feedback, which depends on the acuity of photoreception, as well as the timing and location of gaze shifts. Remapping mechanisms of spatial coordinates of stimuli, as have been demonstrated in the monkey brain, is also showing how the information of fixations are used to maintain memories from past gaze locations. To reinforce and maintain these spatial maps by gaze shifts could be important to provide visual accuracy for manual actions.

Overall, we believe that understanding the precise nature of the multi-modal information processed for motor skills, and how that processing changes with learning, could provide very important guidelines for experimental physiology and anatomy of the BG-TC system.

Chapter 4

Combining modalities of different latencies for optimal motor control: a computational study

Existing computational system models of the basal ganglia are focused on reward prediction [67, 119, 152], action selection [14, 9], or sometimes both [37, 123, 145]. However, almost all of them are uni-modal (see however [123, 155, 60]). The fact that many modalities project to the basal ganglia [2, 114] has, so far, been less considered by modelers. The modular organization of basal ganglia-thalamocortical (BG-TC) loops and the consistent projections across the striatum suggest that the basal ganglia perform generic computations in parallel of multiple modalities [62], and seems to be the most feasible site for multi-modal sensory integration, although the integration mechanism is not well understood [89].

Further, most actor-critic computational models of the basal ganglia have been event-by-event and discrete, neglecting the requirements of real-time control of motor responses (an exception of note is [98]). For the actor-critic model to be a feasible biological model, it must work for continuous streams of inputs and outputs, and deal with signals with different delays. In particular, the long latencies associated with visuomotor feedback [90, 23, 112] may cause problems for controlling real time movements and for coordinating multiple feedbacks.

In this chapter, we propose a real-time actor-critic model of the basal ganglia, for learning and controlling visuomotor skills. Our framework consists of a critic and of multiple actors, where each actor corresponds to a modality or submodality. We propose that the feedback latency constrains the utility of each actor. Further, we propose that the reinforcement learning algorithm plays a critical role in gating between inputs - only the better feedback signals, presumably those with shorter latency, would be reinforced. Thus, modular inputs are, once learned, gated implicitly by their latency. In our framework, the gating is realized by combination of population coded outputs, sharpened by a softmax function in favour of the module with highest confidence. This mechanism is different from explicit gating [73, 61], where explicit signals are computed to weight the influence of modules on the combined output.

To test our hypotheses and study multimodal interactions, we implemented two motor tasks with different agents - “Experiment I” and “Experiment II”. In both experiments, many combinations of hypothetical latencies of modules are trained until convergence. We then look at the performance and compare the outputs of modalities. In Experiment I, we studied a very simple system to clearly understand the effect of feedback delays. Two modules, which are identical except for their feedback delays, were trained until convergence for a simple arm reaching task. In Experiment II, we studied the interaction between vision and somatosensation in a sequential reaching task. Here, we assumed that a “somatosensory module”, corresponding to a motor skill, is learned under the assistance of a “visual module”, a pre-acquired, general but suboptimal controller guided by visual feedback.

We begin with a brief review on findings of visuomotor delays in humans and primates, before we present the general architecture and learning algorithm (for details on the actor-critic algorithm, see Section 2.2). Then the tasks of the experiments are described, followed by the results of Experiment I & II. We conclude with a discussion in context to other work in theoretical motor control and experimental findings.

4.1. Related work: feedback delays in visomotor control

In contrast to many control systems of machines, where feedback may be instantaneous, delays in biological, visuomotor feedback loops are significant, up to 100-200 ms [90, 23]. In monkeys, the response latencies of visual areas of the brain have been measured relative to the onset of a visual stimuli [146] (see Figure 4.1A). As one might expect, the serial nature of the visual pathways are reflected by the mean response latencies. Similarly, the response latencies of motor, cortical areas with respect to initiation of a reaching movement and its associated visual cue onset time were measured [84] (see Figure 4.1B). Note that the primary motor cortex (M1) responds earlier than parietal area 5 (PA5), although PA5 is a part of the somatosensory motor cortex [83, 82], representing planned, potential movements [4].

The Smith predictor [151] is a control system tailored to overcome the problem of feedback delay in applied engineering. Chris Miall and colleagues asked the question if the Smith predictor may play a role in human motor control, and hypothesized that the cerebellum may implement two instances of Smith predictors [113]. However, a recent experiment from the same group did not support Smith prediction as a plausible control scheme for manual tracking [111]. Also, it has been shown that unstable plant dynamics will also destablize the Smith Predictor controller [109].

Mehta & Schaal [109] compared the performance of several control architectures for a delayed, linear pole-balancing model. Behavioural experiments ruled out most architectures, since the observed control gains were too high for these to be stable. The study favoured a forward model in the pre-processing stage of the control loop, as subjects could cope with a long blank-out (no visual feedback) period.

In general, the issue of sensory feedback delay in biological motor control has been relatively little investigated [111]. Several behavioural studies report degraded performance for artificial delays 50 ms and greater [92, 93, 51, 129], but results on delays in motor control beyond accuracy dependence in the literature are scarce.

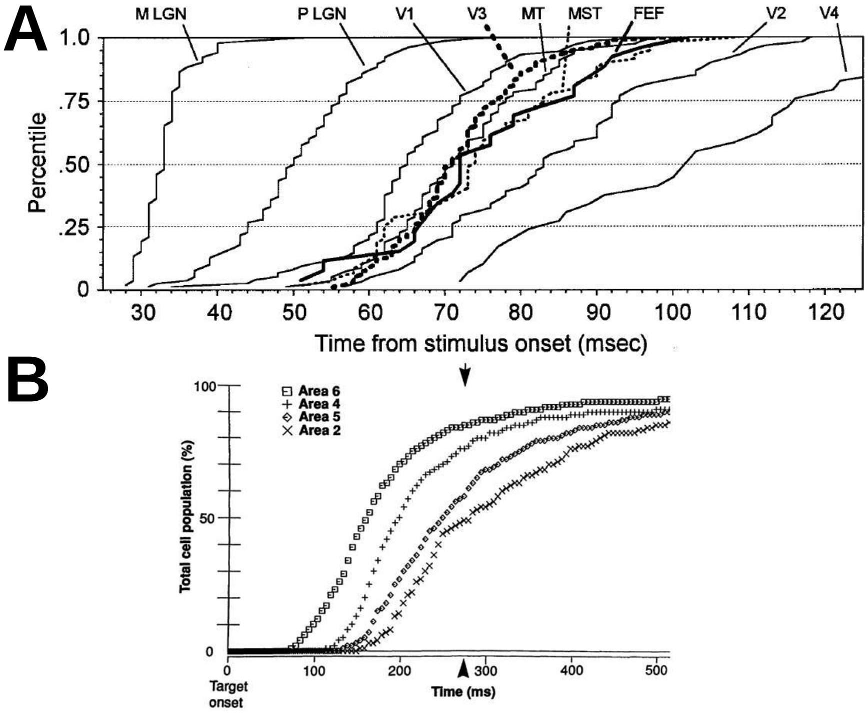

Figure 4.1. Cumulative distributions of population latencies measured in macaques. (A) Response latencies to a visual onset across the visual system, in anesthetized monkeys. M, magnocellular, P, parvocellular, LGN, lateral geniculate nucleus, V1-4, visual areas 1-4, MT, middle temporal area, MST, medial superior temporal area, FEF, Frontal eye field. From *Schmolesky et al.* [146]. (B) Response latencies of neural populations in the motor system relative to onset of a visual cue (0 ms) and initiation time (mean) of an associated reaching movement (arrow at ∼280 ms). Area 6, dorsal premotor cortex (PMd), area 4, primary motor cortex (M1), area 5, parietal area 5, area 2, primary somatosensory cortex. From *Kalaska & Crammond* [84]

4.2. A general framework

As a simple model of learning control using multiple delayed feedback channels, we consider a modular architecture as shown in Figure 4.2. The state $\mathbf{x}(t)$ of the physical environment evolves depending on the motor command $\mathbf{u}(t)$. The state is monitored through different sensory channels $\mathbf{y}^m(t)$ with different delays τ_m $(m = 1, ..., M)$. Each module outputs a population-coded motor command $a^m(t)$, and through their combination $\pi(t)$, the final motor command $u(t)$ is sent out to the physical environment. The goal of control is to maximize the cumulative reward $r(t)$, as in the standard reinforcement learning paradigm [10, 38] .

Below we outline the operation of the feedback control modules, combination of their outputs, and the learning algorithm. The architecture presented here is a modification of a previous report [18] .

Feedback control modules

Each module m has a characteristic feedback signal

$$\mathbf{y}^m(t) = \mathrm{f}^m(\mathbf{x}(t - \tau^m)) \tag{4.1}$$

where $\mathrm{f}^m()$ is an observation function and τ^m is a characteristic latency for the particular module. Each module gives as output a population code

$$\mathbf{a}^m(t) = \mathrm{g}(\mathbf{y}^m(t); \mathbf{w}^m) \tag{4.2}$$

where $\mathrm{g}(t)$ is a function approximator, with a set of trainable parameters $\mathbf{w}^m$. Each element a_j^m $(j = 1, 2, ..J)$ corresponds to a prefered motor output $\bar{\mathbf{u}}_j$.

Combination of modular outputs

The motor command $\mathbf{u} \in R^D$ is represented by a combination of the population coded outputs of all modules with a softmax function:

$$\pi_j(t) = \frac{\exp\Big(\beta \sum_{m=1}^{M} a_j^m(t) + n_j(t)\Big)}{\sum_{j=1}^{J} \exp\Big(\beta \sum_{m=1}^{M} a_j^m(t) + n_j(t)\Big)} \tag{4.3}$$

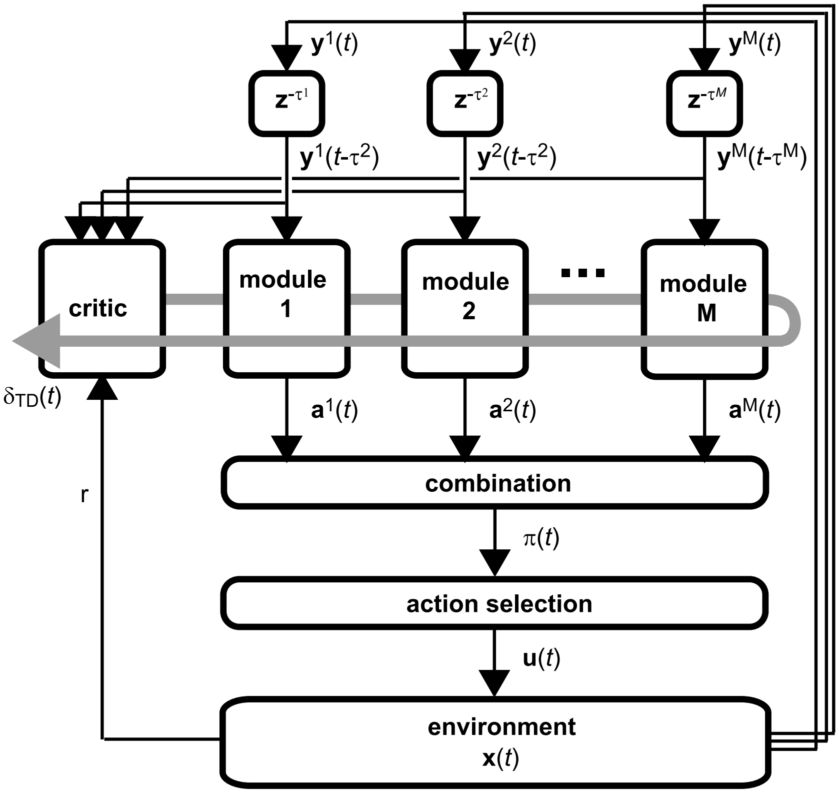

Figure 4.2. The modular network architecture for learning control with multiple feedback pathways (see text).

where β is a constant that regulated the overlap of population codes. The noise term $n_j(t)$ makes the policy stochastic, i.e. it controls the exploration of the agent. The actual motor command $\mathbf{u}(t)$ is given by the weighted sum of the preferred motor commands $\bar{\mathbf{u}}^j$ corresponding to each population code:

$$\mathbf{u}(t) = \sum_{j=1}^{J} \pi_j(t)\bar{\mathbf{u}}^j. \tag{4.4}$$

The modular outputs a_j^m can be interpreted as the log-probability of selecting the output $\bar{\mathbf{u}}_j$,

$$a_j^m(t) = \log(P(\bar{\mathbf{u}}_j(t)|\mathbf{y}^m(t-\tau^m), \mathbf{w}^m)). \tag{4.5}$$

Summing over all modules, adding noise and exponentiating gives the full probability $P(\bar{\mathbf{u}}_j(t)) = \pi_j(t)$. This simple but straightforward interpretation gives a direct relationship between the activities of single neurons and distributional population codes [137, 172] .

Actor-critic learning

Our model implements a form of the continuous actor-critic [38] . The learning algorithm is described in Section 2.2. In the present implementation, the critic learns to estimate the value function from available feedback:

$$V(\mathbf{y}^1(t), \mathbf{y}^2(t), .., \mathbf{y}^M(t); \mathbf{w}^c) \tag{4.6}$$

where $\mathbf{w}^c$ is a set of trainable parameters.

4.3. Implementations of visuomotor reaching tasks

We test the effects of different sensory feedback delays in two simulated experiments of arm reaching. In Experiment I, we used two somatosensory feedback control modules with different delays for a simple reaching task. The aim is to see how the minimal delay affects the control performance and how relative feedback delay affects the selection of the modules by learning. In Experiment II, we used both visual and somatosensory feedback modules for a sequential reaching task. The aim is to see whether and how transition from slow, task independent visual control to fast, task dependent somatosensory control happens under different feedback delays.

Figures 4.3 and 4.6 show the implementation of Experiments 1 and 2, respectively.

Reaching tasks

We use a 2DOF arm, where each link is 0.3 m long, 0.1 m in diameter, and 1 kg (see Figure 4.3). The state is defined by its shoulder and elbow joint angles θ_1 and θ_2 and angular velocities $\dot{\theta}_1$ and $\dot{\theta}_2$. The Cartesian hand position is $\boldsymbol{\xi}^{hand}(\theta_1, \theta_2)$

. The arm moves according to the motor command $\mathbf{u}(t) = (u_1(t), u_2(t))$. In Experiment I, we assume the system noise proportional to the motor command so that each joint torque is given by:

$$u_d^{actual}(t) = (1 + n_d(t))u_d(t) \qquad (d = 1, 2) \tag{4.7}$$

where $n_d(t)$ is white noise with unit variance and mean zero. In Experiment II, we assumed the system noise to be zero.

In Experiment I, the goal is to move the hand as quickly and accurately as possible to the target position T given the start position S. The reward signal is given by an exponential function of the distance of the hand to the target

$$r(t) = a\exp(-b||\boldsymbol{\xi}^{hand}(t) - \boldsymbol{\xi}^{target}||) + c \tag{4.8}$$

where $a = 6$, $b = 20$ and $c = -0.3$. Each trial lasts for 1.0 second.

In Experiment II, the task is to press three targets in consecutive order, which always appear one at the time at the same positions, marked 1, 2 and 3 in Figure 4.6. A target is pressed when the hand reaches a proximity of the target $||\boldsymbol{\xi}^{hand}(t) - \boldsymbol{\xi}^{target}|| < \xi^{prox}$ at a low speed $||\dot{\boldsymbol{\xi}}^{hand}(t)|| < v^{prox}$) ($\xi^{prox} = 0.02$ m and $v^{prox} = 0.5$ m/s). After each successful target reaching, the agent is rewarded with an increasing amount (50, 100, and 150) and the next target appears immediately. Each trial ends after successful completion of the sequence, or after 5 seconds.

Feedback control modules

In Experiment I, we use two somatosensory feedback controllers, while in Experiment II, we use somatosensory and visual feedback controllers.

The somatosensory module The somatosensory control module uses a population code representing joint angles $\boldsymbol{\theta}$ and angular velocities $\dot{\boldsymbol{\theta}}$ of the arm as the input:

$$y_k^m(t) = \frac{1}{Z}\exp\Big(-\frac{1}{2}\Big\{\sum_d(\frac{\theta_d(t-\tau_m) - \bar{\theta}_{kd}}{\sigma_{kd}})^2 + \sum_d(\frac{\dot{\theta}_d(t-\tau_m) - \bar{\omega}_{id}}{\sigma'_{kd}})^2\Big\}\Big) \tag{4.9}$$

where $k = 1, 2, ..K$ is the index of the input units, $\bar{\theta}_{kd}$ and $\bar{\omega}_{kd}$ are their preferred joint angles and velocities, σ and σ' are their width parameters, and Z is a normalization term. In Experiment II, we introduced additional units representing the time since the target onset.

The output of module m is given by another population code

$$\mathbf{a}^m(t) = \mathbf{W}^m \mathbf{y}^m(t - \tau^m) \tag{4.10}$$

where $\mathbf{W}^m$ are trainable weight matrices. Initially all weights are zero. See appendix C for more details.

The visual module The input for the visual feedback controller is the Cartesian positions of the hand and the target

$$\mathbf{y}^v(t) = \{\tilde{\boldsymbol{\xi}}^{hand}(t), \boldsymbol{\xi}^{target}(t - \tau^v))\}. \tag{4.11}$$

While the target position is subject to feedback delay τ^v, we assume that an estimate of the present hand position $\tilde{\boldsymbol{\xi}}^{hand}$ is available, e.g., by simple linear prediction. The output is expressed as a population code $\mathbf{a}^v$.

We assume that the feedback control of the visual module (indexed by v) is pre-acquired and use a linear feedback controller with inverse dynamics compensation and output smoothing (see appendix D). The controller produces bell-shaped velocity profile similar to natural hand movement.

Actor-critic learning

To promote effective exploration, we use a low-pass filtered noise $\mathbf{n}$ in the action output (Equation 4.3)

$$\tau^n \dot{\mathbf{n}}(t) = -\mathbf{n}(t) + \nu \mathbf{N}(t) \tag{4.12}$$

where the time constant $\tau^n = 50$ ms and $N(t)$ is Gaussian noise with zero mean and unit variance. The amplitude ν is fixed at 0.1 in Experiment I, and reduced as $1/(1 + 0.0001T)$ at trial T in Experiment II.

The critic takes the population coded somatosensory feedback as its input and use a linear weighting to produce the value estimate

$$V(t) = \mathbf{w}^c \mathbf{y}(t), \tag{4.13}$$

where $\mathbf{y}(t) = (\mathbf{y}^1(t), \mathbf{y}^2(t))$ in Experiment I and $\mathbf{y}(t) = \mathbf{y}^s(t)$ in Experiment II. (We verified that inclusion of visual input $\mathbf{y}^v$ in Experiment II did not affect the results in this example).

Performance measures

In order to compare the control performance under different settings of feedback delays, we use a number of performance measures, namely, the hand trajectory, the hand velocity profile, the cumulative reward, and the performance time.

The cumulative reward is given by

$$R = \int_0^T r(t)dt, \tag{4.14}$$

where T is the length of a trial (T=1 sec in Experiment I, T=5 sec in Experiment II).

To compare the relative contribution of different modules, we define the actor weight ratio, the output deviation, and the relative output proximity. The actor weight ratio (AWR) we define as the ratio of the absolute sum of actor weights of respective trained module:

$$\mathrm{AWR} = \frac{\sum_k \sum_j |w^1_{jk}|}{\sum_k \sum_j |w^2_{jk}|}, \tag{4.15}$$

i.e. a value AWR > 1 indicates a relatively more influential actor of module 1. We also define the output deviation

$$\mathrm{d^m}(\mathbf{x}(\mathrm{t})) = \mathbf{u}(\mathbf{x}(\mathrm{t})) - \mathbf{u}^{\mathrm{m}}(\mathbf{x}(\mathrm{t})) \ \ (\mathrm{m} = \mathrm{v}, \mathrm{s}) \tag{4.16}$$

which shows how much module m's output differ from the agent's at time t. Its time average over trajectories is given by

$$< \mathrm{d^m} >= \frac{1}{\mathrm{T}} \int_0^{\mathrm{T}} |\mathbf{d}^{\mathrm{m}}(\mathrm{t})| \mathrm{dt}. \tag{4.17}$$

for a trial terminating at T. From the output deviation, we define the relative output proximity

$$\begin{aligned} p^v(t) &= 1 - \frac{d^v(t)}{d^v(t) + d^s(t)} && (4.18) \\ p^s(t) &= 1 - p^v(t) && (4.19) \end{aligned}$$

which is a measure of the relative influence of visual and somatosensory modules over agent output, respectively. Both $p^v(t)$ and $p^s(t)$ are bounded between 0 and 1 and their relation is $p^v(t) + p^s(t) = 1$ by definition. Thus, the value of one module larger than that of the other module indicates that the former is a dominant module.

Simulation

In both experiments the same learning parameters were used: inverse temperature $\beta = 10$, time constants $\tau^{TD} = 200$ ms, $\tau^{ET} = 200$ ms, and learning rate $\alpha = 0.1$ s^{-1}.

All differential equations were approximated with the Euler forward method with a time step small enough not to affect the results (10 ms).

4.4. Simulation results

4.4.1 Experiment I: Simple reaching with somatosensory feedback modules with different delays

In Experiment I (Figure 4.3), we investigate how feedback latencies affect learning and control in the proposed framework. We use a simple implementation with two modules which are identical, except for their feedback latencies. The task is to learn a simple reaching movement with a 2DOF arm. We train the networks with different pairs of feedback latencies and compare their performance and relative contribution of modules after learning has converged.

Reaching trajectories

Figure 4.4 shows examples of hand trajectories generated by the architecture with four different settings of feedback latencies. The 10 trajectories in the top row (Figure 4.4A) are generated with both modules after 100,000 training trials. Effective reaching movement is achieved by all four latency pairs in a robust manner. In the case of $(\tau_1, \tau_2) = (0, 50)$, the variability is higher than in the other

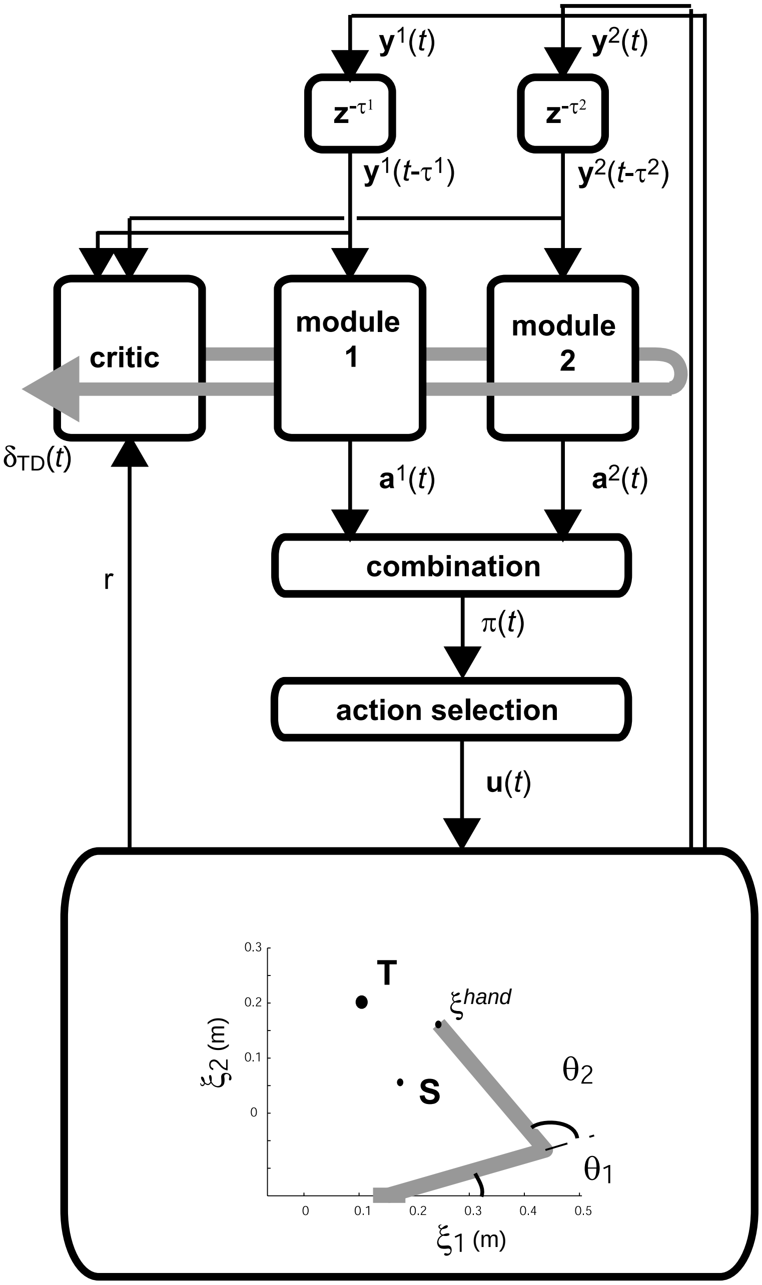

Figure 4.3. The implementation of (τ_1, τ_2) = Experiment I. The agent controls a 2DOF arm by applying joint torques to shoulder and elbow joints, with angles θ_1 and θ_2 respectively (ξ_1 and ξ_2 define the Cartesian coordinates). The task is to reach from the start hand position S to the target position T, as quickly and accurately as possible (according to the reward signal r). The feedback signals $\mathbf{y}^1$ and $\mathbf{y}^2$ are identical, a population code of joint angles and velocities. See text for further details.

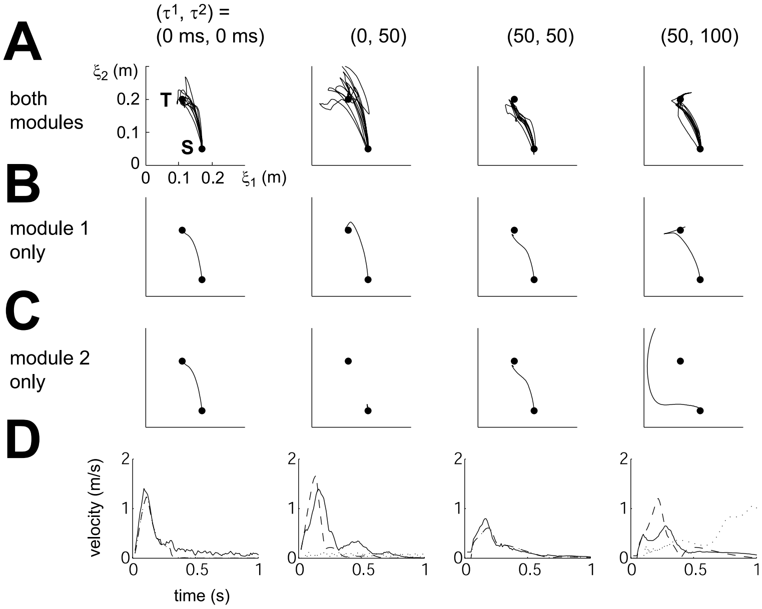

Figure 4.4. Trajectory samples generated by four different settings of latencies (τ_1, τ_2) = (0,0), (0,50), (50,50), and (50-100) milliseconds (ms) after 100,000 trials of training. S - start, T - target. (A) 10 trajectory samples generated by both modules under system noise. (B) System noise free trajectory generated by module 1 only. (C) System noise free trajectory generated by module 2 only. (D) Velocities of both modules (solid line, mean of 10 samples), module 1 only (dashed line) and module 2 (dotted line).

examples. Since the reward function (see Section 4.3) does not explicitly penalize variability, this is still a performance optimally close to other well-performing agents (see below). However, in the cases of (τ_1, τ_2) = (0, 0), and (0, 50) ms the movement is much faster than (50, 50), and (50, 100) ms, as can be seen in the hand velocity plots in the bottom row (Figure 4.4D, solid lines (mean velocity of the samples in Figure 4.4A)). This shows that the shortest feedback delay is critical for the performance. This is not a trivial finding as the output of the module with the longer feedback delay can interfere with the feedback command generated by the module with shorter delay.

In order to see the relative contribution of the two modules, we compared the trajectories generated by either one of the modules (Figure 4.4B-C), with

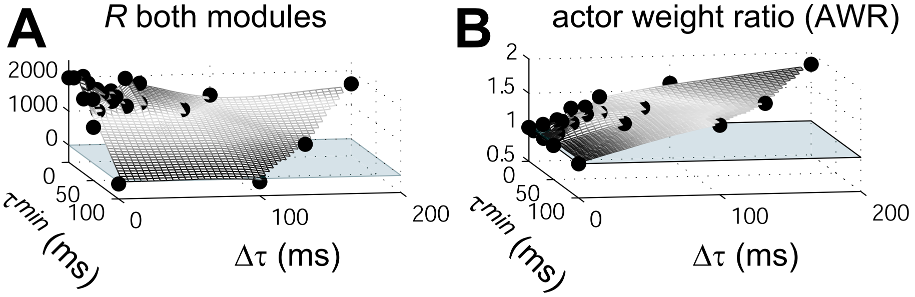

Figure 4.5. Latency-dependent performance measures, displayed as surface plots constrained by 24 latency pairs (black dots) over latency space $(\tau_{min}, \Delta\tau)$. (A) Cumulative reward R, indicating behavioural performance of agent. (B) Actor weigth ratio (AWR), indicating relative contribution of modules. A value below the x-y plane (AWR = 1) indicates relatively larger contribution of slower module 2, above plane larger contribution of faster module 1.

the other module's output set as $\mathbf{a}^m(t) = \mathbf{0}$. With the identical delays (τ_1, τ_2) = (0, 0) and (50, 50) ms, both module can realize comparable trajectories. On the other hand, with different delays (τ_1, τ_2) = (0, 50) and (50, 100) ms, while the module 1 with shorter delay can realize nice trajectories, the module 2 with longer delays generates very poor trajectories. This shows that the less desired outputs of the module with longer feedback delay is effectively shut down by reinforcement learning.

Effects of minimum and relative delays

To verify the critical role of the minimum feedback delay in control performance and the role of relative feedback delay for module selection, we measured the cumulative reward R and the actor weight ratio of trained agents, for 24 different pairs of feedback delays. Under the condition of $\tau_1 \leq \tau_2$, we plot those measures in the parameter space of $\tau_{min} = \tau_1$ and $\Delta\tau = \tau_2 - \tau_1$.

Figure 4.5A shows how the cumulative reward depends on τ^{min} and $\Delta\tau$. Each black dot corresponds to a trained agent with the specific latency pair. It is clearly seen that the longer τ_{min} results in reduced cumulative reward, while the relative delay $\Delta\tau$ has almost no effect on the performance. Figure 4.5B shows the actor weight ratio, which increases markedly with the increase of the relative delay $\Delta\tau$ (for $\tau^{min} = 0$, per definition AWR $\equiv 1$. Never was AWR < 1.).

These results confirm that the performance of the modular learning control architecture is mostly determined by the module with shortest latency. This is achieved by the softmax combination of population coded outputs of modules (4.2) and tuning of modular outputs by actor-critic learning. It is noteworthy that potential problem of slower feedback module contaminating the good output of the faster module has been avoided by this scheme.

4.4.2 Experiment II: Sequential reaching with visual and somatosensory feedbacks

In Experiment II (Figure 4.6), we introduce a more realistic, complex implementation of a visuomotor sequence task. In motor skill acquisition, there is substantial evidence for a shift in cortical activity with experience, from prefrontal areas to motor areas [135, 79, 40, 63, 50]. Analogously, there should be a shift in modalities of feedback subserving these cortical areas; from extrinsic (visual) feedback needed for anticipation and proceduralization of task dynamics to intrinsic (somatosensory) feedback needed for optimization of motor control [123].

Here, we study the transfer between these two systems, a “visual module” and a “somatosensory module”, in a task of reaching a stereotyped sequence of three targets. The visual module relies on a general purpose controller which regulates a single reach to a given visual target. We assume that the module is preacquired and is not be optimized for any particular target sequence. The somatosensory module relies on somatosensory feedback and become optimized for repeated motor sequences. Architectures with different latency pairs τ^v and τ^s were trained for 100,000 trials, after which learning had converged in all cases. We investigate the relative contribution of the somatosensory module for different latency pairs and also compare the robustness against external perturbations of the composite system versus single module control.

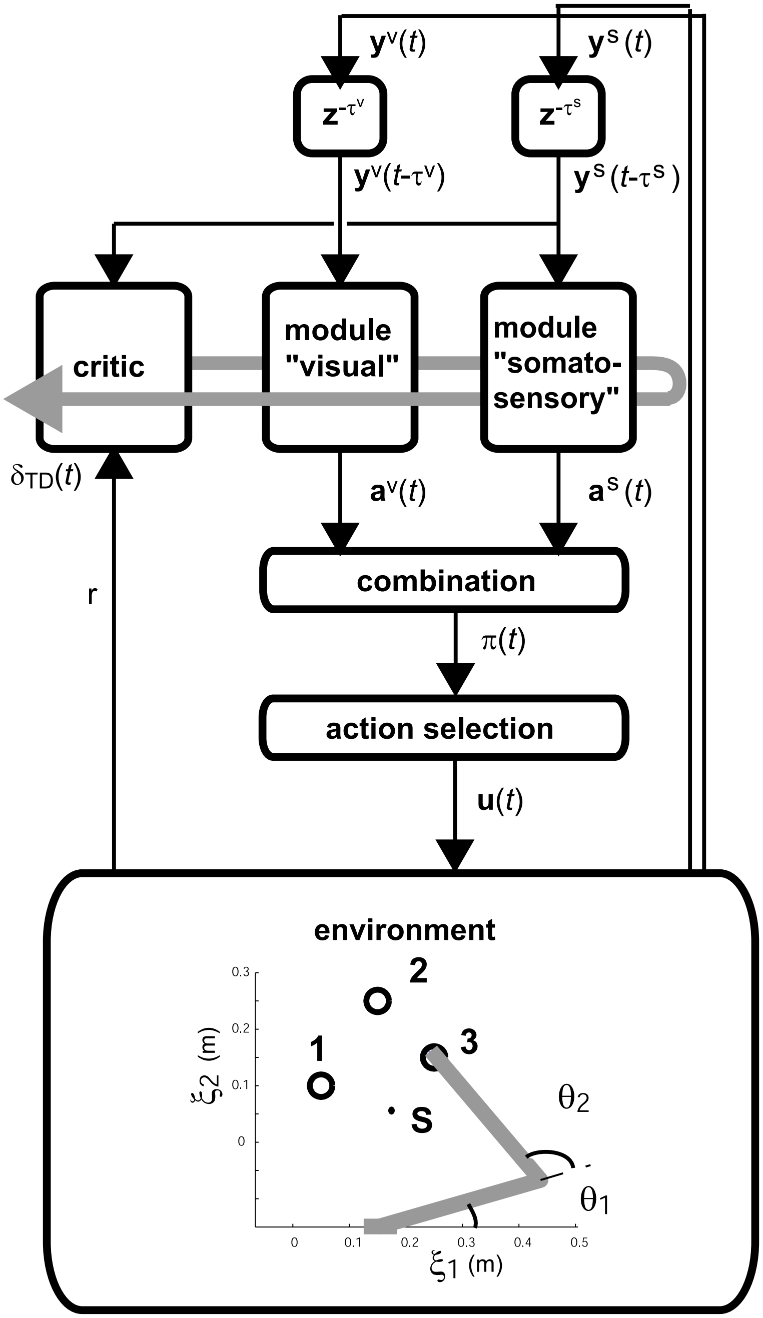

Figure 4.6. The implementation of Experiment II. Here, with the arm as in Experiment I, the goal is to press targets 1, 2 and 3, presented in consequent order, starting from S. Reward is given only at the time when a key is pressed. The agent consists of two modules called "visual module" and "somatosensory module". The visual module is a fixed controller, receiving feedback about the current target position $\boldsymbol{\xi}^{target}$ and hand position $\boldsymbol{\xi}^{hand}$ to control a reaching movement. The somatosensory module is similar to the modules in Experiment I. See text for further details.

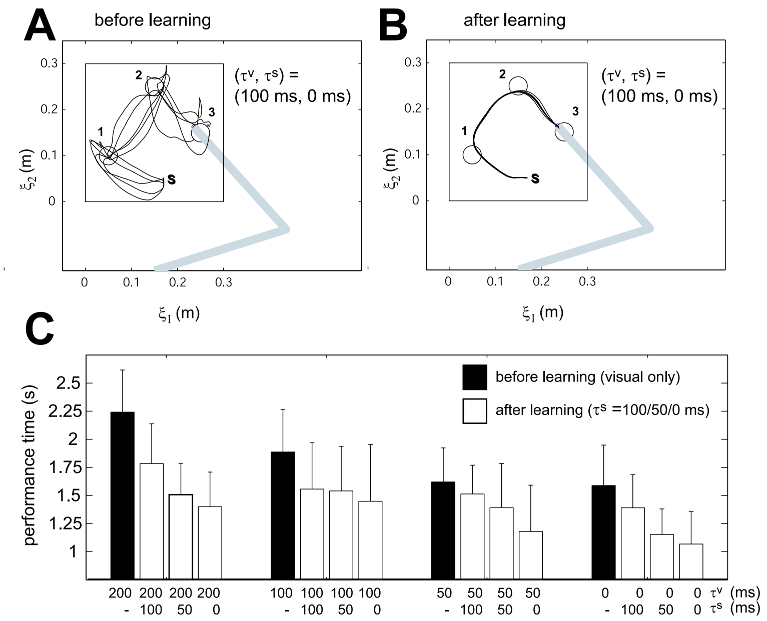

Figure 4.7. Performance before and after learning. (A) 5 sample trajectories before learning. (B) 5 sample trajectories after learning. (C) Performance times of 12 latency pairs before learning (black bars) and after learning (white bars), compared with equal levels of exploratory noise ($\nu = 0.01$). Note that the initial performance is the same for agents with equal τ^v, since the somatosensory module is inactive before learning.

Learning performance

Figure 4.7A-B compare the reaching trajectories before and after learning ($(\tau^v, \tau^s) = (100, 0)$ ms). Before learning, movements are variable and step-by-step - they are directed towards one target at the time. After learning, movements are stereotyped, and also coarticulated, as they are redirected towards targets 2 and 3 before preceding targets are concluded.

Figure 4.7C compares the performance time (the time it takes to complete one trial) before (black bars) and after (white bars) 100,000 trials of learning for 12 different latency pairs. Clearly, sequence-specific learning by the somatosensory module contributes to reduction of the performance time. Its potential to do so is primarily constrained by τ^s, which has a decreasing trend of performance time

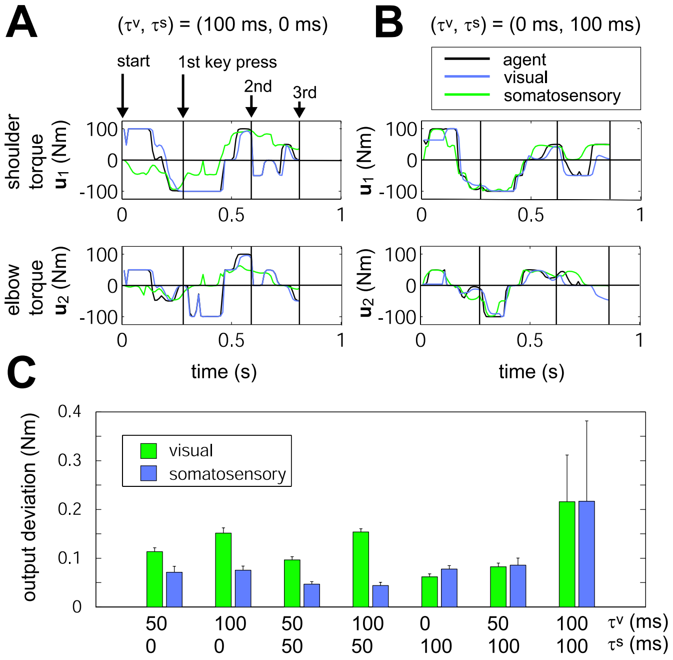

Figure 4.8. A comparison of contribution to joint torque outputs between the visual and somatosensory modules. (A-B) Example trajectories of shoulder (top) and elbow (bottom) torques over time for the latency pairs (τ^v, τ^m)=(100, 0) (A) and (0, 100) ms (B). The green, blue and black lines correspond to the outputs of the visual module, somatosensory module and agent, respectively. (C) A comparison of mean output deviation (100 trials, noise amplitude $\nu = 0.02$) of visual and somatosensory modules, for 7 latency pairs.

for 100, 50 and 0 ms with any latency τ^v of the visual module. In turn, τ^v is also a constraint for performance, as the performance times of learned modules are shorter with lower τ^v.

Contribution of the somatosensory module

To elucidate the contribution of the somatosensory module, we compared the joint torque outputs of single modules (computed as in Experiment I) with the joint torque output of the agent. Figure 4.8A-B shows trajectories of generated joint torques over time (one trial), for the two extremes of relative latency in our study, $((\tau^v, \tau^s) = (100, 0)$ ms (A), $\Delta\tau = 100$ ms) and $((\tau^v, \tau^s) = (0, 100)$ ms, $\Delta\tau = -100$ ms) (B). In the first latency pair, the somatosensory module generates an output different from the visual module, but is evidently dominant as it is close to the agent output. In the second latency pair, the outputs of the two modules are close to each other, indicating that both modules equally contribute to the agent output. Figure 4.8C shows the quantitative picture, expressed as mean output deviation for 7 latency pairs. In cases of $\tau^v < \tau^m$, the visual module has the smaller output deviation ((50, 100) and (0, 100)), indicating larger contribution. In the case of mutual, long latency (100, 100), contribution is equal. Otherwise, the somatosensory module has lower output deviation. This result indicates that for the somatosensory module to learn an independent policy, it needs to have a shorter or equal latency τ^s relative to τ^v, i.e. $\tau^s \leq \tau^v$.

We then investigated how the learned behaviour is driven by the somatosensory module. We compared the normal behaviour of the learned agent with a condition with the visual module inactive. Figure 4.9A shows examples of hand trajectories for four latency pairs in the two conditions. With both modules, all agents are always successful. When the visual module is inactive, the ability to control the movement depends on the relative latency $\Delta\tau = \tau^v - \tau^s$. The success rate of the somatosensory module to complete a trial (given 100 trials) is shown in Figure 4.9B. We observe that the successful rate is high in the case of $\tau^s < \tau^v$, whereas none (or single trials in the case of (50, 50)) was successful in the case of $\tau^s \geq \tau^v$. These results further confirm our observation above (Figure 4.8) that the somatosensory module can become dominant as far as $\tau^s \leq \tau^v$.

Figure 4.9C compares the mean performance time of the two conditions. On average, two of the agents ((50, 0) and (100, 50)) can on average perform almost as well in the somatosensory only condition, but note the smaller variance of the normal condition. The visual module provides robustness also late in learning.

Robustness to perturbation

To further evaluate the robustness of the composite system, we perturbed a behaving agent $((\tau^v, \tau^s) = (100, 0)$ ms) by applying a force on the end effector

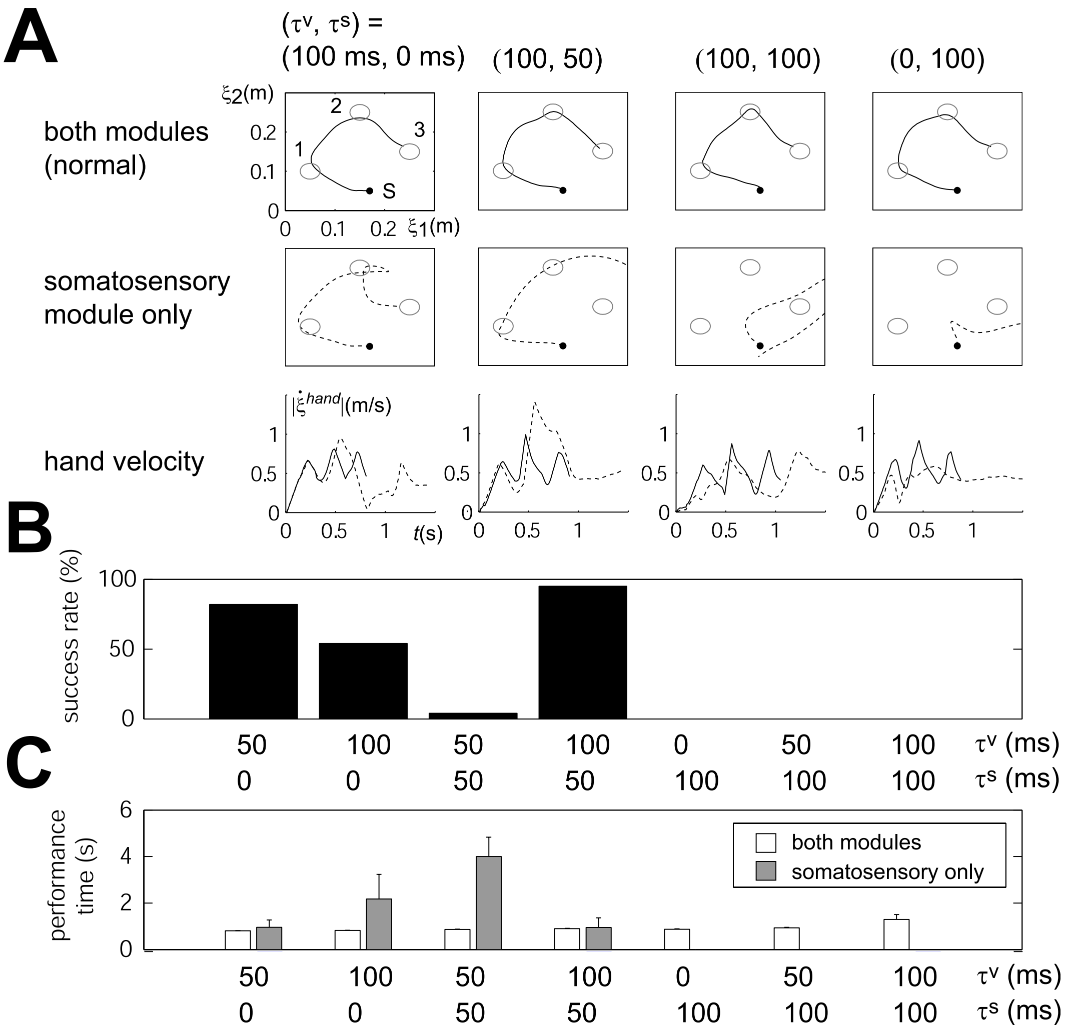

Figure 4.9. (A) Learned behaviour of the agent in normal execution ("both modules", top row, solid lines) and in execution with the visual module inactive ("somatosensory module only", middle row, dashed lines) for four different latency pairs (noise-free). The bottom row shows corresponding, absolute hand velocities (first 1.5 seconds) over time. (B) Success rates of the "somatosensory module only" condition for 7 different latency pairs ($\nu = 0.02$). For the agents of $\tau^s = 100$, there were no successful trials in this condition. (C) Comparison of performance times between normal (white bars) and somatosensory module only (gray bars) conditions, for successful trials.

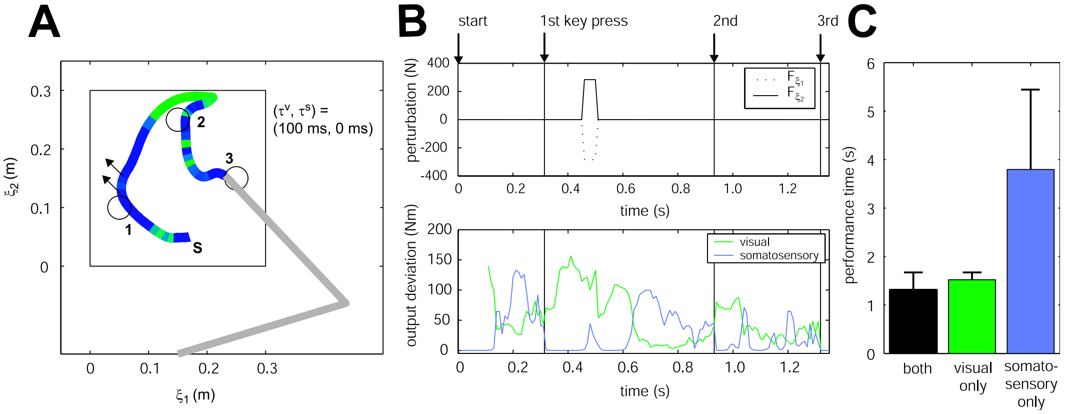

Figure 4.10. External force perturbation imposed on the end effector (hand), for a trained agent $(\tau^v, \tau^s) = (100, 0)$ ms. (A-B) Example trajectory when an impulse (400 N, 50 ms) perturbs the composite system. (A) Spatial movement trajectory. The two arrows indicate the direction of the force perturbation, drawn from the position of start and stop of the impulse. The green/blue colour indicate relative proximity to visual/somatosensory modules' output, respectively, i.e. green indicates $p^v(t) > p^s(t)$ and blue $p^v(t) < p^s(t)$. (B) Temporal trajectories (corresponding to (A)) of impulse (top), and output deviations of visual (green) and somatosensory (blue) modules. Note that output deviation is inverse proportional to proximity in (A). (C) Mean PT for perturbed trials with an impulse (400 N, 50 ms) of random direction and random onset (0.3-0.6 s from trial start), for the agent $(\tau^v, \tau^s) = (100, 0)$ ms, comparing control with both versus single module.

.

(hand). An impulse with constant force (400 N) was applied for 50 ms in a random direction (in the plane of the arm), for 50 ms, 0.3-0.6 s after trial start. Figure 4.10A-B shows an example trajectory, where the impulse (400 N up left, onset at 0.3 s) throws the agent off track to miss target 2 to the left. The green/blue colours of the trajectory indicate the relative proximity (see Section 4.3) in Figure 4.10A of visual/somatosensory modules' output to the agent's, respectively. Note how the visual module predominates after the perturbation, to put back the trained movement on track (towards target 2), after which the somatosensory module regains influence anew. Figure 4.10C shows a comparison of the impact on performance time (mean of 1000 trials) for the random impulse, when operating with both or single modules active. The visual module functions as a safeguard against perturbations, since the somatosensory module alone (blue bar) cannot effectively recover, resulting in the significantly higher performance time (for which 56 % of the trials were timed out at 5.0 s). The somatosensory module contributes to speedup before and after perturbation recovery, which is why both modules (black bar) are performing faster than the visual only (green bar).

In summary, these results indicate that in this visuomotor sequence task, as learning progresses, the somatosensory module with the presumably shorter latency becomes dominant in motor control. After learning, the visual module provides stability when the effector ends up outside the well-trained regime. The memory transfer, or the degree of control by different modalities, critically depends on the difference in latencies between the visual and somatosensory modules.

4.5. Discussion

In this chapter, we have examined how feedback latency affects the relative importance of modules for the learning and control of real-time motor skills. With softmax combination of population-coded output of multiple control modules, we demonstrated in simulations how the modules with shorter latency attain dominance in motor control. Although the result may sound straightforward, there are potential problems with conflicts of multiple modules, e.g., the longer latency output pulling back movement by the shorter latency module. It is noteworthy that appropriate module selection was achieved without any explicit gating and simply by reinforcement of the output of the module that best contributed the performance. The last experiment showed that module weighting is highly flexible; the general-purpose visual module takes over the job when the trajectory-specific somatosensory module does not perform well.

4.5.1 State estimation models

In dealing with delayed or noisy sensory signal, a recently popular paradigm is to use recursive Bayesian filters to estimate the hidden state [163, 161, 97] (see Section 2.1.3). Such a model may also explain more weight on the faster module that is more informative about the current state. However, Bayesian inference requires the models of the physical dynamics and sensory delay and noise and also takes heavy on-line computation, except for linear Gaussian systems where Kalman filtering is possible. Instead, here we pursued much simpler, model-free approach of training feedback controllers specialized for given delays. Analysis of pros and cons of these approaches and their possible integration is the subject of our future study.

4.5.2 Motor skill learning

The mechanism of transfer from declarative to procedural memories is poorly understood [39, 65] . In our framework, modules with shorter latency become dominant with learning. As demonstrated in experiment II, this allows specialized motor skills based on fast, intrinsic feedback loops to emerge under general purpose controllers based on slow, extrinsic feedback like vision or audition. If the difference in feedback latency is long enough, the faster modality will eventually become independent of the slower modality, which can then be used for other purposes.

There are two analogies between our framework and the BG-TC system: 1) its organization into modular circuits [2], and 2) the actor-critic architecture [67]. In previous experimental [62, 65] and computational [123] work, we have proposed that prefrontal and motor BG-TC loops cooperate in motor sequence learning, encoding sequences in visual and motor coordinates, respectively.

In experiment II, faster movements were learned even though reward was given only for key presses, regardless of time expenditure. Rewards received faster are valued higher because of temporal discounting of rewards (Equations 2.2.2 & 2.2.2. This property may naturally explain why performance of numerous skill learning tasks (e.g. [5]) speeds up, although speed is not an explicit performance criterion. Brain mechanisms of reward discounting is an active research topic [108, 149, 155, 33].

4.5.3 Multimodal integration

In our framework, the critic was used for evaluation of the combined output. This implies that each loop knows the combined output, for learning to be possible. In the simpler model by Nakahara et al. [123], each loop evaluated its action separately, with coordination at the perceptual (input) level. Such implementation is consistent with the fact that BG-TC loops do not converge anatomically. It is not understood how BG-TC loops are coordinated, whether actions are evaluated locally like a multi-agent system, or globally, which is a much simpler credit assignment problem. Integration may be possible by recurrent cortical networks, or by spiral connections between the striatum and substantia nigra [58, 60].

4.6. Conclusion

The success of this rather simple modular learning control framework motivates future studies with agents comprising of more complex, heterogeneous features, such as different sensor noise levels, learning speeds, or inclusion of feedforward components. For example, given a slow, low-noise module and a fast, noisy module, the former would be used for precision tasks and the latter for speed tasks. To further test the generality of this prediction, delayed auditory feedback could be added as a third modality, and modality dependence could be tested under different pairs of feedback delays.

A challenge for this model to be interpreted as a basal ganglia-thalamocortical model is that there is no evidence for integration of loop outputs. For our model to be plausible, recurrent cortical connections must mediate a global signal conveying information about the collective action. Further anatomical, electrophysiological and imaging studies need to be done to understand how information is conveyed between loops.

The brain receives possibly thousands of sensory signals, from which it has to make a sensible response. Biological reinforcement learning may not just be about selecting actions, but also about selecting sensory input. In this context, feedback latencies may be a critical factor for which input and output connections are formed.

Appendix

4.A. Population codes

In both experiments the population codes were equal. In the somatosensory modules, the preferred joint angles $\bar{\theta}_{kd}$ and angular velocities $\bar{\omega}_{kd}$ were distributed uniformly in a 7 x 7 x 3 x 3 grid ($K_0 = 441$ nodes) for $k = 1, 2, ..K_0$ nodes , in the ranges (-0.2:1.2,1,2:1.6) rad and (-1:1,-1:1) rad/s. The corresponding variances σ_{kd} and σ'_{kd} were half the distance to the closest node in each direction.

The preferred joint torques $\bar{\mathbf{u}}_j$ corresponding to action j were distributed symmetrically over the origin in a 5 x 5 grid, in the range (-100:100,-100:100) Nm with the middle (0,0) unit removed. The corresponding variances σ''_{jd} were half the distance to the closest node in each direction.

The somatosensory module in experiment II also included “context units”. The context units consists of 3 tapped delay lines, each corresponding to a key in the sequence task. Each delay line had 8 units, i.e. 24 context units in all. For the k-th unit in the n-th delay line ($k > K_0$, $k \neq K_0 + 8(n-1) + 1$):

$$\dot{y}_k^m(t) = -\frac{1}{\tau^C} y_k^m(t) + y_{k-1}(t) \tag{4.20}$$

where $\tau^C = 30$ ms. Each delay line is initiated by the input at ($k = K_0 + 8(n - 1) + 1$):

$$y_k^m(t) = \delta(t - \tau_n^{keypress}) \tag{4.21}$$

where δ is the Dirac delta function, and $\tau_n^{keypress}$ is the instant the n-th key was pressed.

4.B. The visual controller in Experiment II

The feedback signal $\mathbf{y}^v$ to the visual module consists of the hand kinematics $\boldsymbol{\xi}^{hand}$, $\dot{\boldsymbol{\xi}}^{hand}$ and the target position $\boldsymbol{\xi}^{target}$. Since the computed torque control law itself does require at least a good estimation of the current motor kinematics, the delayed feedback signals will not produce satisfactory control: the delays will cause oscillations, and become unstable at some 50-100 ms. To overcome this problem, we assumed that the agent has a good model of its own internal dynamics, and can cancel out the delay of $\boldsymbol{\xi}^{hand}$ with a prediction $\tilde{\boldsymbol{\xi}}^{hand}(t) = \boldsymbol{\xi}^{hand}(t)$. The target position $\boldsymbol{\xi}^{target}$ is assumed not to be predictable. Thus, with the onset of a new target, it takes τ^v ms before the visual module reacts towards that target. The control is further perturbed by a decoding error, by modification of the somatosensory module and by the stochasticity of action selec tion.

The joint torques are first computed by

$$\dot{\mathbf{u}}^{visual}(t) = -\frac{1}{\tau^{CT}}\mathbf{u}^{visual}(t) + \lambda \mathbf{u}^{visual\prime}(\ddot{\tilde{\boldsymbol{\xi}}}^{hand}, \dot{\tilde{\boldsymbol{\xi}}}^{hand}, \mathbf{e}) \tag{4.22}$$

where τ^{CT} and λ are constants, $\mathbf{e} = \boldsymbol{\xi}^{target}(t - \tau^v) - \tilde{\boldsymbol{\xi}}^{hand}(t)$ and the input to the filter is the inverse dynamics equation

$$\mathbf{u}^{visual\prime}(t) = \mathbf{J}^T(\mathbf{M}(\dot{\tilde{\boldsymbol{\xi}}}^{hand} + \mathbf{K}_1\dot{\tilde{\boldsymbol{\xi}}}^{hand} - \mathbf{K}_2\mathbf{e}) + \mathbf{C}\dot{\tilde{\boldsymbol{\xi}}}^{hand}) \tag{4.23}$$

in Cartesian coordinates, where $\mathbf{J}$ is the Jacobian $(\partial\boldsymbol{\theta}/\partial\tilde{\boldsymbol{\xi}}^{hand})$, $\mathbf{M}$ the moment of inertia matrix and $\mathbf{C}$ the Coriolis matrix. Using a filter by Equation 4.22, more bell-shaped velocity profiles of the hand, similar to biological motion are generated, in contrast to using Equation 4.23 directly.

The module output is an expansion of the joint torque $\mathbf{u}^{visual}$ on a population vector

$$a_j^v(t) = \frac{1}{Z}\exp(-\frac{1}{2}\{\sum_d(\frac{u_d^{visual}(t) - \bar{u}_{jd}}{\sigma''_{jd}})^2\}) \tag{4.24}$$

where Z is the normalization term, $\bar{u}_{jd}$ is a preferable joint torque for Cartesian dimension d for vector element j, σ''_{jd} the corresponding variance.

The parameters of equations of 4.22 and 4.23 were $\tau^{CT} = 50$ ms, $\lambda = 100$, $\mathbf{K}_1 = [10\ 0;0\ 10]$, $\mathbf{K}_2 = [50\ 0;0\ 50]$.

Chapter 5

Learning gaze strategies for control of sequential hand movement: an experimental study

Like most animals with advanced vision, human eyes are mobile. This is advantageous for several reasons. First of all, it makes us able to stabilize the field of view, as a response to self-induced perturbations: head rotations or locomotion. Secondly, the higher acuity of the fovea makes it favourable to refixate towards objects of particular interest. Thirdly, it allows us to track moving targets in the environment. In most human behaviours, we see how gaze is refixating (or sometimes pursuing) actively. Seen as a constraint, a natural question to ask is which location of gaze fixation is better or worse at any given time. By studying gaze behaviour we can gain important insights for how visual feedback is used for motor control.

In this chapter, we focus on the role of gaze in manual skills. Our aim is to study the change of gaze as subjects improve on a motor task over a long time. We design a task, “the 1 x 20 task” (an adoption of “the m x n task” [64, 62] , see Section 3.1.4), where subjects learn to press a stereotype sequence of key presses on a touch screen. With extensive training over five days, subjects gradually learn to execute automatically with virtually no reaction time. We then analyze the spatiotemporal properties of gaze trajectories in early and late training. Gaze strategies are likely to change with learning for three reasons. First, changes in trajectories of the hand and manipulated objects require an updated gaze strategy. Second, there may be a limit on how fast gaze can be shifted for efficient feedback, which forces economizing of gaze shifts if task events occur at high frequencies. Third, accurate prediction of task variables allows subjects to optimize gaze for visual feedback. Alternatively, learning of proprioceptive and tactile feedback and feedforward control would make execution independent of vision, and gaze does not need to respond to the task.

This chapter is organized as follows. First, we describe the experimental paradigm of the sequential reaching task, including setup, measurements and analysis. Then we propose a Bayesian model of gaze-dependent updating of spatial representation is presented. Then we present the results of the experiment. Finally, we discuss what significance our results have for visuomotor research, and ask questions we find important for future studies.

5.1. Methods

5.1.1 Experimental design

Eight healthy, right-handed subjects (7 male and 1 female; ages 22-35) with normal (or corrected to normal by contact lenses) vision were used in this study. Subjects gave informed consent and the experiments were approved by the institutional ethics committee.

Setup

Subjects were seated in a dentist chair with their shoulders restrained in a harness (Figure 5.1A). The subjects were able to freely move their right arm, but the setup restricted movement of their torso. Located directly in front of the subject within easy reach of their right arm was a touch screen (Elo 1925L, Elo touchsystems; 19 in., 1280 x 1024 pixels, 60 Hz) on which the subject's task was presented. The timing and location of the subject's finger presses were recorded for later analysis (response time of the touch screen: 10-15 ms).

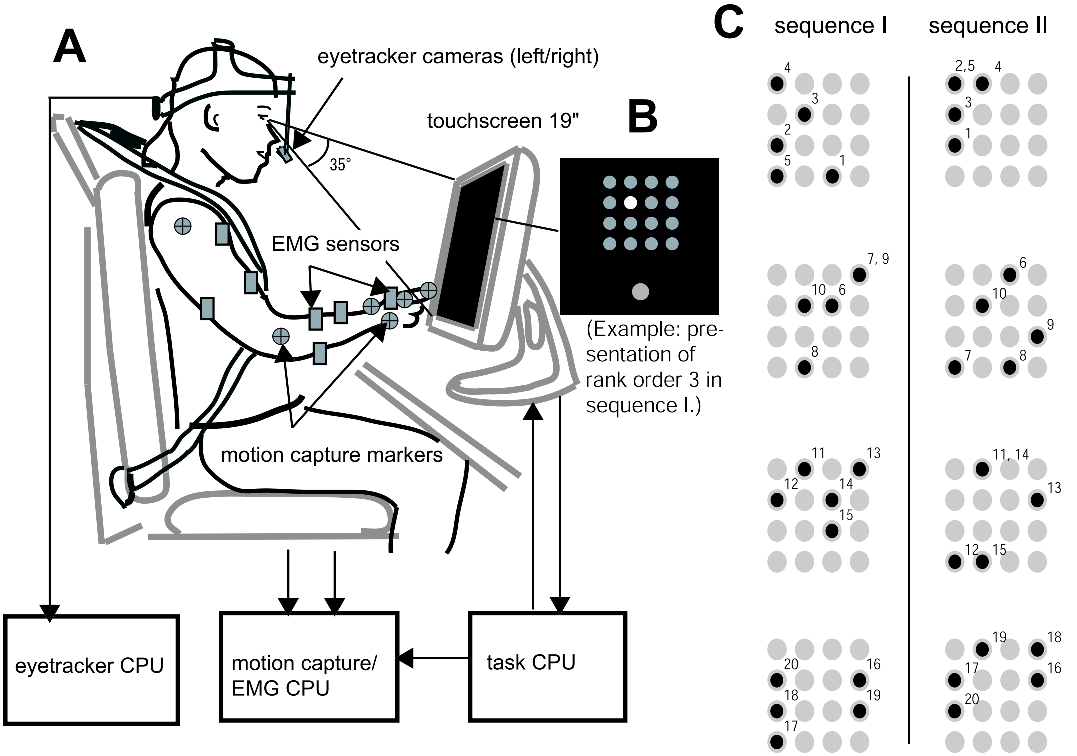

Figure 5.1. Experimental setup. (A) The subject was seated in a dentist chair in front of a touch screen. A racing harness was used to limit movement of the torso. Gaze from both eyes, motion capture (OPTOTRAK), and EMG (first and last day of training only) from the right arm were recorded. (B) Screenshot. Target buttons were arranged in a 4 x 4 target matrix at the top of the screen, below which was the home button. At any one time there was a single active target, displayed in white. (C) The two sequences used in the study, displayed in rank orders (numbers) five at a time (segmentation is irrelevant and only for visualization here; in the task, targets are presented one at a time, as in Figure B).

Eye gaze data was recorded using the EyeLink II system (SR Research Ottawa). The gaze location for each eye onto the touch screen was recorded at 250 Hz.

Motion capture of six locations on the right arm was recorded simultaneously. On the first and last day of the experiment, electromyography from seven muscles was also recorded. The results from these measurements will be reported elsewhere.

Task paradigm

To study long term behavioural changes in motor sequence/skill learning, subjects were asked to perform a motor sequence task, adopted from the m x n task [64, 62] . For a given trial, subjects were to press a stereotype sequence of 20 targets. Targets were displayed as circular buttons (24 mm diam.) in a 4 x 4 *target matrix* (indexed by $i = 1, 2, .., 16$), 39 mm between button centers in vertical and horizontal direction (Figure 5.1B). A single home button (34 mm diam.) was presented 67 mm below this matrix. All buttons were drawn in a gray colour on a black background, except for the current, active target button, which was displayed in white. Subjects controlled the start of a trial by pressing the home button, after which the first target was presented. There were no explicit errors or timing constraints in the task; a target was presented until the subject pressed it, invoking the next target. All targets were onset with a 100 ms delay after the preceding press. Subjects were instructed to complete the sequence as fast as possible, with voluntary pauses between trials.

Subjects experienced two sequences of 20 targets on 5 consecutive days. On a single training day, “sequence I” (see below) were presented 100 times (equivalently 100 trials), and then “sequence II” 100 times. Thus, each sequence was performed 500 times by each subject during the experiment. The analysis is mainly focused on the it success trials (error-free trials) as they could be compared across subjects and days (see below).

Sequences

We refer to the ordinal positions of a target in the sequences as its *rank order* or *rank order target*, which are indexed by $j = 1, 2, .., 20$. In analysis, we also sort contexts by rank movement time and rank movement length (see below). Sequences I & II were generated randomly (see Figure 5.1C), with one constraint: the hand would never occlude the consequent target.

5.1.2 Analysis

Sequence geometry

For each rank order j , there is an associated target i at position $\mathbf{x}_{i(j)} = \left\{x_{i(j)}, y_{i(j)}\right\}$, where elements are horizontal and vertical Cartesian screen coordinates, respectively. There is also an associated preceding *movement* j, by which we mean the displacement (of the index finger) from the previous target center $\mathbf{x}_{i(j-1)}$ to the present target center $\mathbf{x}_{i(j)}$. Movements are characterized by *length* (Euclidean distance) and *direction* (angle from the x axis). We also define two geometrical concepts including more than two targets: the *relative angle* is the angle between two consequent movements, and the *center of mass* (COM) is the mean position of n consequent targets $[j, j+1, .., j+n-1]$ with respect to the current rank order j:

$$\mathbf{x}_{COM}\left(j,n\right) = \frac{1}{n}\sum_{m=j}^{j+n-1}\mathbf{x}_m \tag{5.1}$$

where n is the number of targets included. Targets are always equally weighted.

Performance measures

By *performance time* we mean the time it takes to complete one trial, which is the primary performance measure, what the subject is asked to minimize. By *movement time* we mean the time it takes to execute a single movement, i.e. the time elapsed between two button presses.

We are also interested in accuracy. We label touches on the screen outside the active target as *errors*. A trial with no errors is referred to as a *success trial* or an *error-free trial*. We distinguish one particular kind of error which we call *misses*, while *other* is used for other errors. *Misses* are defined as touches in the near vicinity of the target ($<$ 15 mm from the key perimeter, where the subject clearly intended to press the illuminated target. *Other* includes for example miss-directed reaches and next-key-in-sequence-hits (when a correct button press was not recognized by the touch screen).

To investigate how sequential context influence accuracy, we define the *relative miss frequency* (RMF) to be the relative frequency of misses of a particular rank order. For rank order$j$,

$$\text{RMF}(j) = (\text{number of misses at } j)/\ (\text{number of misses in sequence})$$

given a set of trials. This measure allows for direct comparison between subjects with varying levels of accuracy to determine the difficulty of success of a particular rank order. If not stated otherwise, the RMF is averaged across all subjects and training days.

Representation of time

Since a trial consists of many events that are not controlled in time, it is not always convenient to compare entire trials in absolute time. For easy comparison, we introduce a *normalized event time*, defined as

$$T(t, j) = j + \frac{t - t_j}{t_{j+1} - t_j} \qquad t_j \leq t \leq t_{j+1}. \tag{5.2}$$

That is, time is indexed by the last pressed key with rank order j, and to that the fraction of time elapsed to the next key press is added.

Measurement of gaze

Gaze was recorded from both eyes. The data from the left eye was used exclusively as a measure of gaze; the right eye was used for verification. While the Eyelink system compensates for head movement, subjects were instructed to avoid it as much as possible. In the results, we generally report on *gaze shifts*, and their associated timings and locations are reported by the start time and the screen coordinate of the left eye fixation following the saccade. Saccade and blink detection was defined the same as for the default psychophysical configuration of the Eyelink II system (thresholds for saccade detection (velocity and acceleration): 22 deg/s, 4000 deg/s^2). Fixations lasting less than 100 ms were omitted from the event data in the results reported.

A calibration procedure was conducted between every 10 trials, where subjects fixated 9 consecutive targets for 1.5 seconds each, appearing on random locations. The standard deviations of the residuals over (8 subjects) x (5 days) x (9 points) x (9 calibrations) were 11 mm (horizontal) and 17 mm (vertical) (a 1.0 degree eye rotation corresponded to 8.8 mm on the screen). Furthermore, it was necessary to perform an additional drift correction between each trial. This was done using the assumption that the mean position of gaze and hand would be equal, allowing a correction for the deviation of the gaze from the hand. The standard deviations were 16 mm (horizontal) and 24 mm (vertical).

Subjects S7 and S8 had considerable noise in the gaze measurement due to oscillation of the cameras. However, the signals were recovered using a zero phase, digital Butterworth notch filter (8 Hz to 24 Hz, 12th order). (The spatial data of subject S8, day 5 was too poor to be included in the analyses.)

To synchronize the Eyelink system with the task computer and other measurements, we exploited behavioural data in two steps. At the start of each training session and sequence, subjects tracked a target oscillating horizontally on the screen (at 0.1 Hz, 7 cm amplitude) with both gaze and index finger, for 3 cycles. We then used this temporal “fingerprint” on each data set to synchronize them. To further improve accuracy, we took advantage of the 9 x 9 fixations in the custom calibrations, where targets were unpredictable. We assumed that the average latencies from target onset to saccade initiation were constant between training sessions for a single subject, and that the median of the estimated latencies of all sessions per subject was close to the true latency. We estimated the standard error to 12-40 ms, depending on the subject.

Analysis of gaze

In single trials in the present task, timing of the several movements and key presses are not controlled by the task, but by the subject. There may be several candidate targets for gaze fixation at any given time. Therefore, rather than by timing, gaze must be interpreted with respect to the context that they occur, and by the duration of fixations. For this reason, we categorized fixations by the number N_{kpf} (number of key presses per fixation).

Shorter fixations may be associated with a single target. For fixations ($N_{kpf} = 0, 1$), we associated each fixation to a candidate target, to study timing and location. For each fixation initiated when target j as active ($t_{j-1} < t \leq t_j$) , we classified the fixation to belong to the nearest neighbour in the space of candidate targets $(j-1)$ ("postdictive"), j ("reactive") and $(j+1)$ ("predictive"). If targets $(j-1)$ and $(j+1)$ had the same positions, target $(j+1)$ was assigned. The fixation initiation time was compared to onset and press of the associated target button. All fixation locations were compared in a relative coordinate system $\{u, v\}$, where u and v are the lateral and axial deviations from the associated target, with respect to the preceding movement.

As for gaze fixations ($N_{kpf} \geq 2$) , the association with a single target may be less relevant. For a gaze fixation initiated when target j was active, the spatial location of the fixation was assigned to the nearest of $\mathbf{x}_{COM}(j-1, N_{kpf})$ ("postdictive"), $\mathbf{x}_{COM}(j, N_{kpf})$ ("reactive") and $\mathbf{x}_{COM}(j+1, N_{kpf})$ ("predictive"). Then, the spatial location of the gaze fixation was compared to the associated COM, and two other candidate hypotheses: "closest target" (distance to the closest of targets included in COM) and "first target" (distance to first target included in COM).

Statistical sampling

Most of the analysis is limited to success (error-free) trials. To compare "early" and "late" learning, we use the data of day 1 and day 5, respectively. Since subjects have different number of success trials, we sometimes use only the last 9 success trials of day 1 and 5, to weigh subjects equally.

5.1.3 A model of dynamic updating of spatial representation

In this section, we propose a model that addresses how spatial accuracy is constrained by gaze. We assume that the brain maintains a representation of target positions $\mathbf{x}_i (i = 1, 2, .., 16)$ in Cartesian (screen) coordinates, independent of body, head, and eye movement. We do not consider depth.

The general idea is simple: the spatial representation is continuously updated by integration of visual feedback, improving the certainty of target position where the subject is currently fixating, while it is deteriorating elsewhere. The integration is done within a simple Bayesian framework.

Assume that each key position $\mathbf{x}_i$ is represented in the brain by a probability distribution $p(\mathbf{x} = \mathbf{x}_i)$, which we denote as $p(\mathbf{x}_i)$. Drift causes its variance $\sigma_i^2(t)$ to diffuse by a constant velocity ν (we assume that the distribution is radially symmetric, and can thus be described by a scalar variance). With visual input $s(t - \tau^{delay})$ (delayed in the neural pathways by τ^{delay}), the estimate is sharpened by multiplication of the likelihood $p\left(s\left(t - \tau^{delay}\right)|\mathbf{x}_i\left(t\right)\right)$ and the prior $p\left(\mathbf{x}_i\left(t\right)|\mathbf{x}_i\left(t - \Delta t\right)\right)$, which is the distribution of the previous time step $(t - \Delta t)$, subject to diffusion. Then, the update of the posterior is

$$p(\mathbf{x}_i(t)) \propto p(s(t - \tau_{delay})|\mathbf{x}_i(t)) \int p(\mathbf{x}_i(t)|\mathbf{x}_i(t - \Delta t)) p(\mathbf{x}_i(t - \Delta t)) d\mathbf{x}_i(t - \Delta t). \quad (5.3)$$

The initial prior $p(\mathbf{x}_i\left(t = 0\right))$ is the non-visual estimate of the position. If we assume that the distributions are all Gaussian, centered at the true position $\mathbf{x}_i$, the update of the variance σ_i^2 becomes

$$\frac{1}{\sigma_i^2(t)} = \frac{1}{\sigma_s^2(t - \tau^{delay})} + \frac{1}{\sigma_i^2(t) + \nu \Delta t} \quad (5.4)$$

where $\sigma_i^{-2}(t)$ is the inverse variance or the *certainty* about target i. The variance $\sigma_s^2(t)$ of the likelihood $p\left(s|\mathbf{x}_i\right)$ depends on the eccentricity of the target. Its inverse is also assumed to be a Gaussian:

$$\frac{1}{\sigma_s^2(t)} = \begin{cases} 0 & \text{during saccades,} \\ A_0 \exp\Big(-\frac{1}{2}\Big\{\big(\frac{x_i - x^{gaze}(t)}{\sigma^{gaze}}\big)^2 + \big(\frac{y_i - y^{gaze}(t)}{\sigma^{gaze}}\big)^2\Big\}\Big) & \text{otherwise} \end{cases} \quad (5.5)$$

where $\mathbf{x}^{gaze}$ is the gaze location, and A_0 is the acuity for targets with no eccentricity. The parameter σ^{gaze} determines the relative scale of the eccentric acuity for foveal and peripheral vision.

The quantity σ_i^{-2} is a measure of the certainty of the estimate of the position $\mathbf{x}_i$ of the i-th target. The value of σ_i^{-2} approaches 0 when the target is not in view. Fixation in the vicinity of the target gradually increases the certainty, which saturates for a fixation durable enough.

The certainty σ_i^{-2} about a target position may be relevant when the movement was planned, continuously during execution or by the end of the movement. In this study, we assume that the end of the movement is the most relevant (see 5.3). For each rank order j , we refer to the value of $\sigma_{i(j)}^{-2}(t = t_j)$ as σ_j^{-2} , where t_j is the time the key was pressed. We use σ_j^{-2} to predict the measured accuracy of movements, in terms of misses.

When not mentioned otherwise, we used a standard assumption of parameter values for calculations: diffusion rate ν 5.0 mm^2/s, feedback delay τ^{delay} 200 ms, acuity 0.01 mm^{-2}, eccentric acuity parameter σ^{gaze} 6.0 deg. (53 mm), and saccade duration D 50 ms.

Model prediction

In the Results section, we evaluate the model by testing the correlation of certainties σ_j^{-2} with relative miss frequencies (RMF). As we expect misses to be a low-probability event, we also expect miss frequencies to correlate with average behaviour, rather than being sensitive to the gaze trajectories of single success trials. To compute σ_j^{-2} , we utilized three different assumptions for gaze trajectory:

1) “target” : a trajectory that has a gaze shift towards the center of each key j, initiated 100 ms before each key press.

2) “mode”: a trajectory based on mode over all success trials. We compute the mode of gaze locations at each press time t_j. For consequent locations where difference in mode is less than 3 deg., we merge into a single longer fixation. Otherwise, we assume a gaze shift of 50 ms duration, initiated 100 ms before the first key press observed.

3) “random”: a trajectory where 20 gaze shifts, each initiated 100 ms before each key press, have random gaze locations over the key panel. This assumption is for control.

To check the robustness of correlation, we perturbed the assumed gaze trajectories 100 times, where each gaze location in "target" and "mode" were perturbed randomly in horizontal and vertical directions (standard deviation: ± 10 mm in each direction). In case of "random", new random locations were generated each time.

We also assume a sequence of key press times t_j , by mean movement times over all success trials.

5.2. Results

The results are composed of four subsections. In the first, we report on the change of performance, in terms of speed and accuracy, with training. In the second, we report on the change in gaze frequency with training. In the third, we investigate the spatiotemporal properties of gaze and its change with training, and look for regularities in timing and location of gaze fixations. Finally, in the fourth, we look for possible relations between performance and gaze.

Performance

Figure 5.2 shows two performance measures over the 5 days of training: performance time (Figure 5.2A) and errors per trial (Figure 5.2B). To determine the significance of trends across trials, we performed an ANCOVA on the performance measures, with a random effects variable of subjects, and a covariate of trial number. Performance times had a significant covariant effect for sequence I ($F_{1,1716} = 4166$, $p < 0.001$) and sequence II ($F_{1,1409} = 3336$, $p < 0.001$), decreasing over trials. For errors per trials, there was a significant increase for sequence I ($F_{1,3991} = 852$, $p < 0.001$) and sequence II ($F_{1,3991} = 181$, $p < 0.001$).

We then investigated performance associated with single movements. Movement times were correlated with movement length, ($R^2 = 0.36$, $p < 0.0001$, for all success trials and all subjects), and weakly correlated with relative angle ($R^2 = 0.059$, $p < 0.0001$). There was no significant correlation to movement direction ($p = 0.55$). We did not observe any chunking pattern in the behavioural data (compare study by [143]).

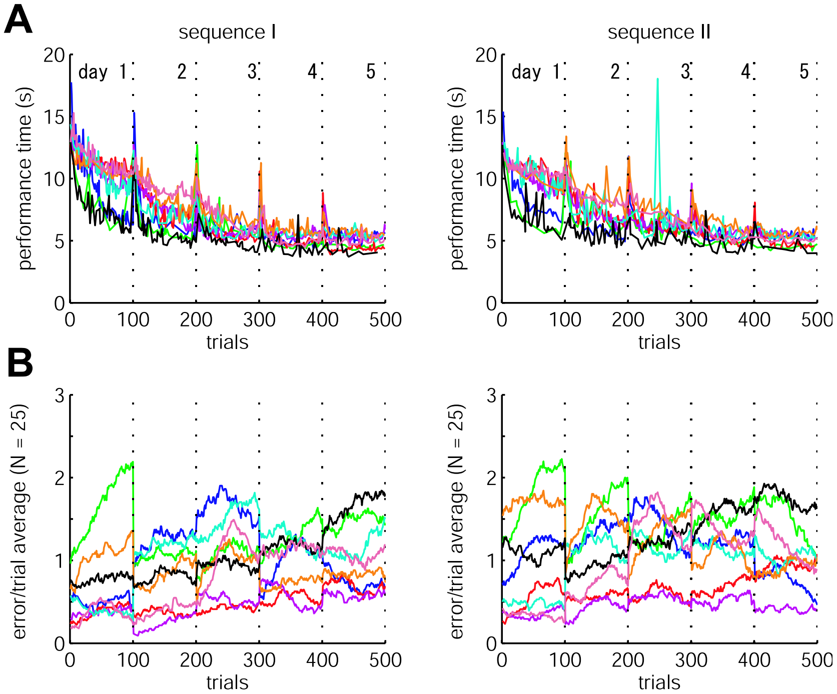

Figure 5.2. Performance time decreases while errors per trial increase over the five days. (A) Performance time for success trials, sequence I (left) and sequence II (right). Each separate day of the experiment is indicated with the dotted vertical line. Each colour corresponds to an individual subject. (B) Number of errors per trial, displayed as a moving average over 25 trials.

The accuracy of hand movements was not uniform across the 20 targets in either of the two sequences. We investigated if the relative miss frequency (RMF) of the population was specific to geometry or rank order. In terms of geometry, the accuracy was weakly correlated ($R^2 = 0.05$, $p < 0.0001$) with movement length, but not to directions of preceding ($p = 0.13$) or succeeding movements ($p = 0.18$), nor to relative angle ($p = 0.08$).

Moreover, we found the RMF to be rank order-specific. Figure 5.3A shows the RMF for each rank order for single subjects and the population mean. We used the student t-test to determine if there was a significant difference between rank order-specific, population mean RMFs, for all possible rank order pairs. For $p < 0.05$, 10 % (79/760 unique pairs) was statistically significant, for $p < 0.005$,

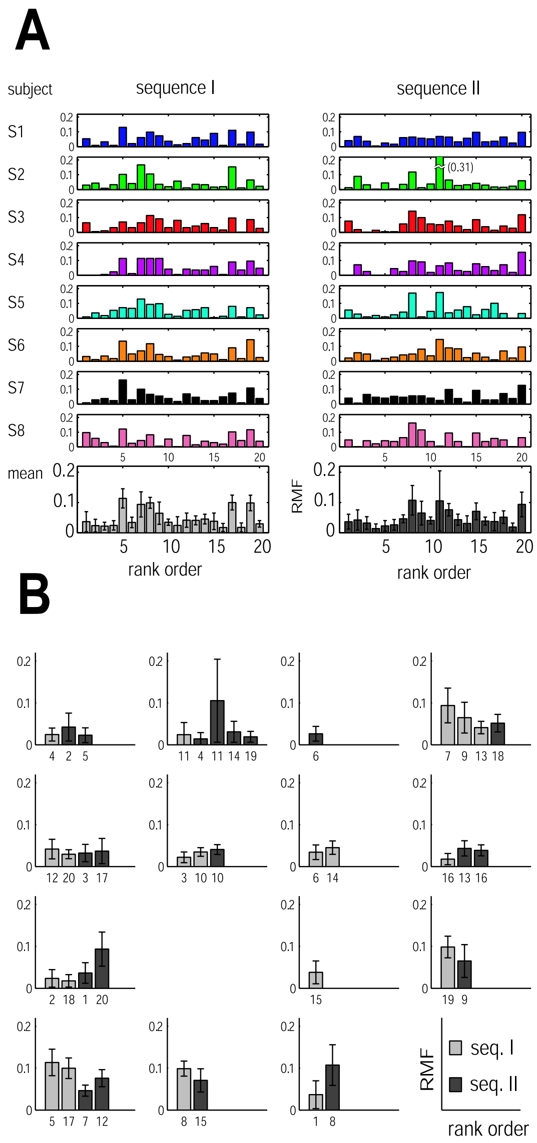

Figure 5.3. Similar patterns of misses are seen across all subjects. (A) Relative Miss Frequencies (RMF) sorted by rank order for single subjects, and for the population mean(bottom panels; sequence I & II in light and dark gray, respectively). Particular key presses tend to be missed more often by all of the subjects. These keys can be seen clearly in the population mean data. (B) The population mean RMFs displayed at their corresponding target positions in the 4 x 4 target matrix. Rank order is given below each bar. The error frequency is not related to the pressing of particular keys. Instead the RMF exhibit clear differences depending on the order in which the key is pressed.

5 % (37/760) (corrected for multiple comparisons (Bonferroni correction)). For correlation between subject RMFs, 85 % of all subject pairs were significant ($p < 0.05$, $R^2 =$ 0.12 (min), 0.59 (max), 0.32 (mean). In Figure 5.3B, the population RMFs are sorted by their associated target position in the target matrix.

The RMFs shown in Figure 5.3 were calculated across all 500 trials. The RMFs were consistent between training days. For example, comparing population mean RMFs for day 5 with days 1-4 gave correlation coefficients 0.35, 0.47, 0.57, and 0.78, respectively.

In summary, all subjects learned to improve the performance over the training period, to execute the sequences 2-3 times faster than on initial trials, with a modest decline in accuracy. The relative accuracy of movements was rank order-specific (context-specific), and not sensitive to general geometrical features, training day or subject.

5.2.1 Frequency of gaze shifts

With experience and faster execution, we observed a reduction of gaze shifts per trial. Figure 5.4 shows how the quantity and timing of gaze shifts (fixation initiation times) changes across training days for one example subject. On day 1, most key presses were preceded by 2 gaze shifts. Then, gradually the number of gaze shifts reduced to 1 or 0 per key press.

Figure 5.5 shows the quantitative change of gaze shifts with experience, for individual subjects: gaze shifts per trial (Figure 5.5A) and gaze shifts per second (Figure 5.5B). We performed an ANCOVA to determine significant trends, analogous to the data in Figure 5.1. For gaze shifts per trial, there was a significant decreasing trend for sequence I ($F_{1,1682} = 3400$, $p < 0.001$) and sequence II ($F_{1,1370} = 2298$, $p < 0.001$). For gaze shift frequency, there was a significant, but weak increase for sequence I ($F_{1,3885} = 991$, $p < 0.001$, $R^2 = 0.11$) and sequence II ($F_{1,3838} = 1831$, $p < 0.001$, $R^2 = 0.02$). On day 5, the average number of fixations per trial was 17.1 ± 2.4 for sequence I and 18.0 ± 2.6 for sequence II.

In summary, we found only a limited increase in gaze shift frequency across the training period, while the frequency of hand movements increased 2-3 times, eventually resulting in fewer fixations than there were targets.

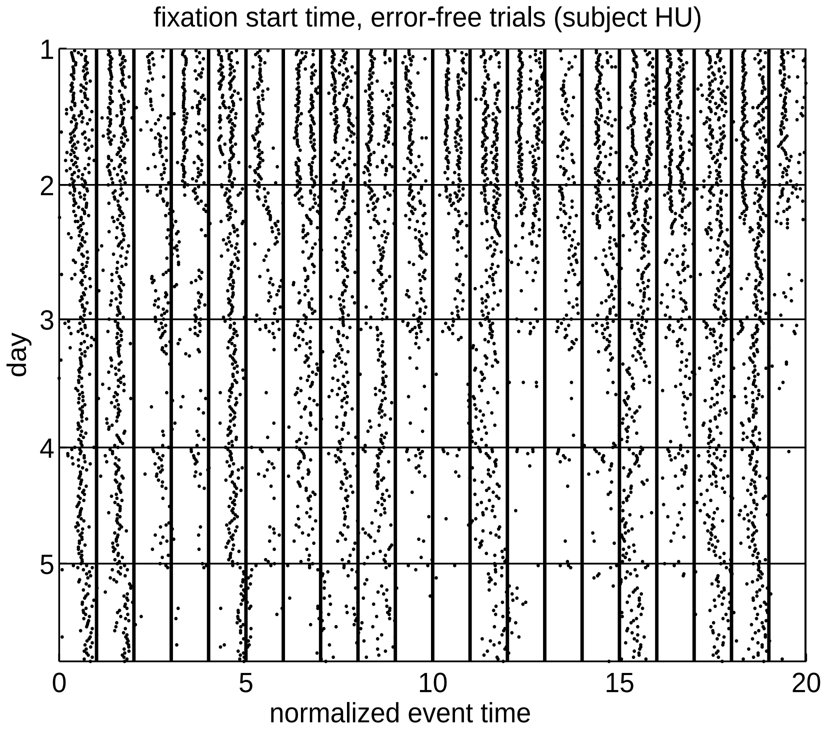

Figure 5.4. The temporal pattern of gaze fixation changes with learning. Temporal pattern of fixation initiation times for subject S3, sequence I are shown for all success trials in chronological order (top to bottom, divided into training days by horizontal lines). Key press times are indicated by vertical black lines, by which time is normalized. Each dot corresponds to a fixation initiation time.

5.2.2 Spatiotemporal properties of gaze

Figure 5.6 shows representative examples of one early (blue lines) and one late (red lines) gaze trajectory. The spatial segments in Figure 5.6A highlight the contexts where early and late behaviour are considerably different. In the early trial, gaze is relocated in response to each new target. In the late trial, single fixations sometimes span several key presses. In this section, we analyze the general timing and location of gaze fixations, categorizing them by the number N_{kpf} of key presses that occurred during the fixation (see Section 5.1). After determining relative occurrence of different N_{kpf} , we first analyze fixations associated with a

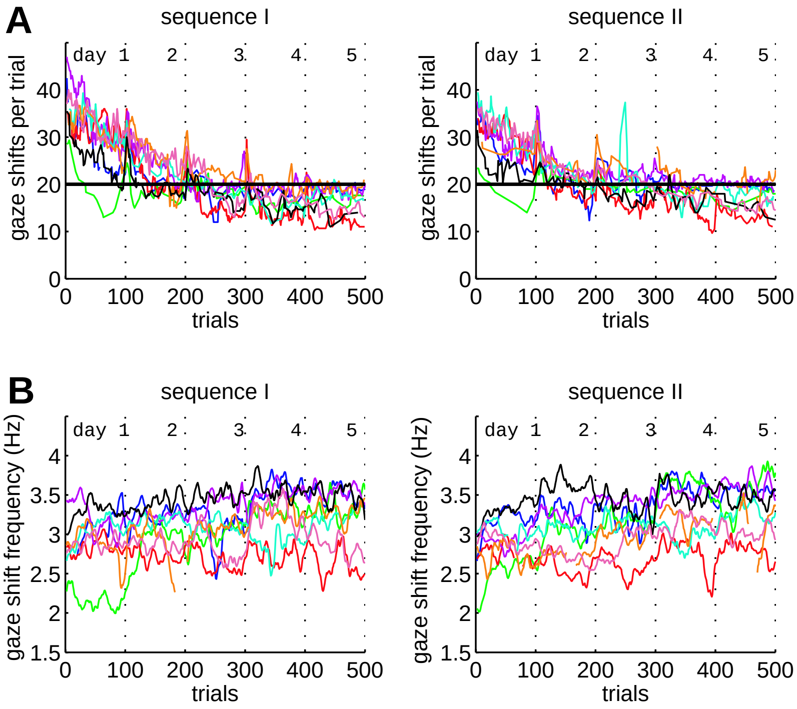

Figure 5.5. The number of gaze shifts per trial is reduced as learning progresses while the frequency of gaze shifts remains steady. (A) Number of gaze shifts per error-free trial. Below the horizontal line at 20 targets, subjects are using fewer fixations than there are targets in a trial. (B) Frequency of gaze shifts for individual subjects. Gaze shift frequencies for all 500 trials (displayed as a moving average of 5 trials), of sequence I (left) and II (right).

single target ($N_{kpf} = 0, 1$), then those that are associated with multiple targets ($N_{kpf} \geq 2$).

Figure 5.7 shows how the distribution of N_{kpf} changed from early (day 1) to late (day 5) training. A single ANOVA (random effect subject, main effects day and N_{kpf}) resulted in a significant interaction effect between N and day ($p < 0.001$), indicating that there was a shift towards higher N_{kpf} e on day 5 relative to day 1. As shown in the figure, paired t-tests of separate N_{kpf} between days indicated significant decrease of $N_{kpf} = 0$, and significant increases of $N_{kpf} = 1$ and $N_{kpf} = 2$.

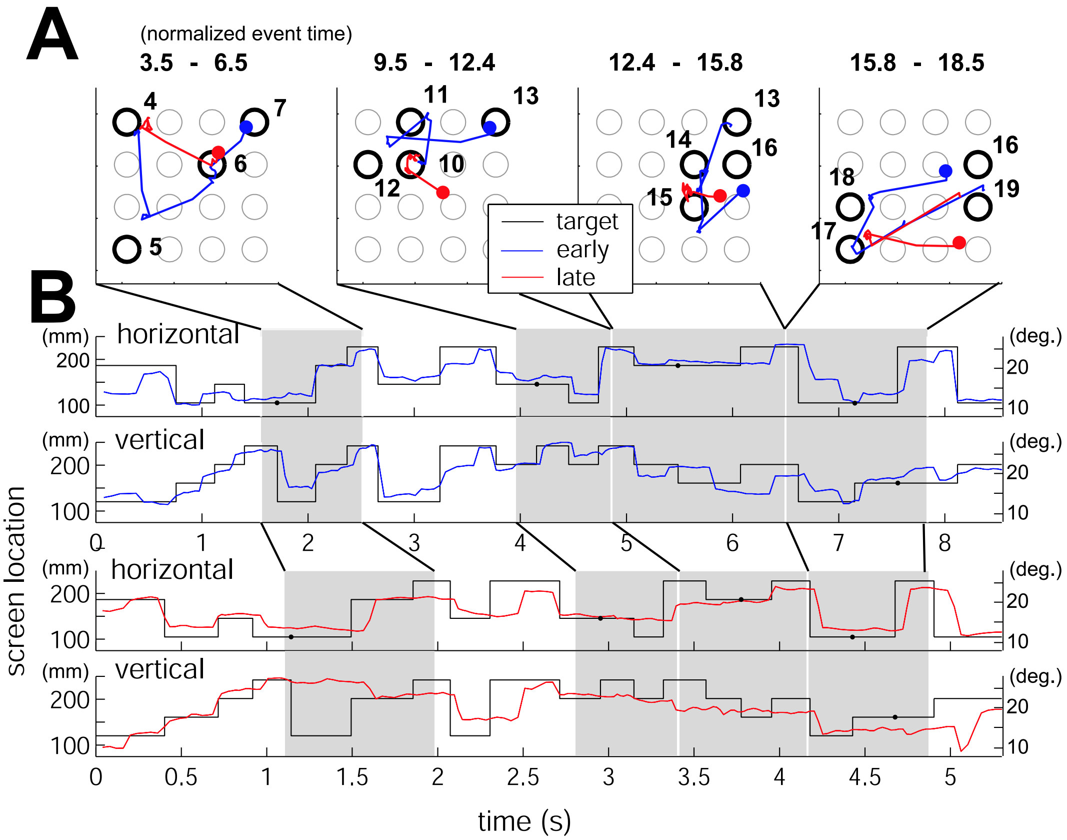

Figure 5.6. The spatial pattern of gaze fixation changes with learning. An example of gaze behaviour for one early (trial 108 (day 2), blue lines) and for one late trial (trial 467 (day 5), red lines) of subject S1, sequence I. (A) spatial 2D plots of temporal segments of trial trajectories, highlighting contexts where early and late behaviour are different. Start and end of segment in normalized event time are shown above. Blue and red dots indicate end of trajectories in respective segment. Target buttons active during the segment are highlighted with bold circles and rank order. (B) temporal profiles of trial trajectories in horizontal and vertical screen coordinates (mm, right ordinates and gaze angle, deg., left ordinates), compared to active target location (black lines). Segments shown in A are shaded. In the early trial (performance time: 8.54 s), the subject ' s gaze responds to each target, in the late trial (performance time: 5.30 s), it often responds to clusters of targets.

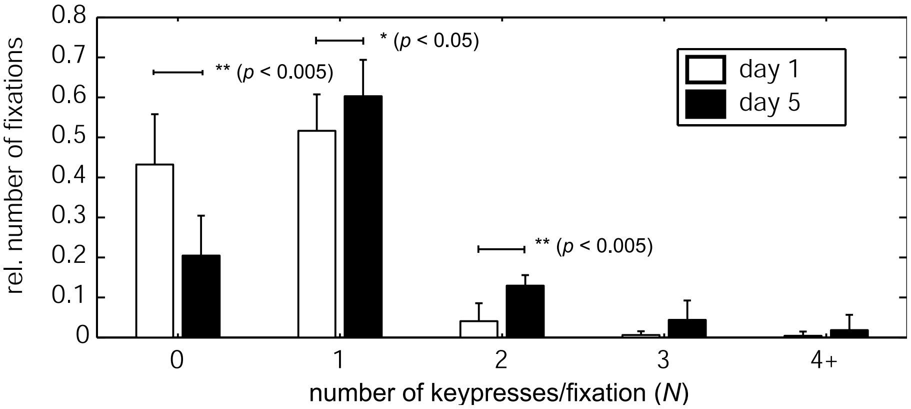

Figure 5.7. The number of key presses per fixation shifts over the learning. The relative frequencies of fixations with $N_{kpf} = 0$, $N_{kpf} = 1$, $N_{kpf} = 2$, $N_{kpf} = 3$ and $N_{kpf} \geq 4$, on day 1 (white bars) and day 5 (black bars) across subjects. The error bars indicate standard deviation. Significant differences between the number of key presses per fixation were tested between days 1 and 5. There was a significant reduction in $N_{kpf} = 0$ and a significant increase in $N_{kpf} = 1$ and $N_{kpf} = 2$ on day 5.

Due to the fast frequency of hand movements (up to 5 Hz), fixations were not always located in vicinity of the active target, but sometimes close to the preceding or following target. To analyze timing and locations of ($N_{kpf} = 0, 1$), we associated each fixation with the closest of preceding ("postdictive"), current ("reactive", and following ("predictive") targets (see Section 5.1).

Figure 5.8 shows the timing of fixation initiations for ($N_{kpf} = 0, 1$) on day 1 and day 5, with respect to key onset (Figure 5.8A/C) and key press (Figure 5.8B/D). The most frequent timing (0.40 s) in Figure 5.8A, we presume is a characteristic reaction time for targets that subjects have not yet learned to predict. In general, the variability of peaks are constrained by (hand) movement times. Consequently, peaks are broader on day 1, and narrower on day 5. A majority of fixations were initiated just before (0.10 s) key presses, both on day 1 and day 5. This short interval from gaze initiation to key press gives the subject little use of foveal feedback - an indication that peripheral vision is used for estimation of target location.

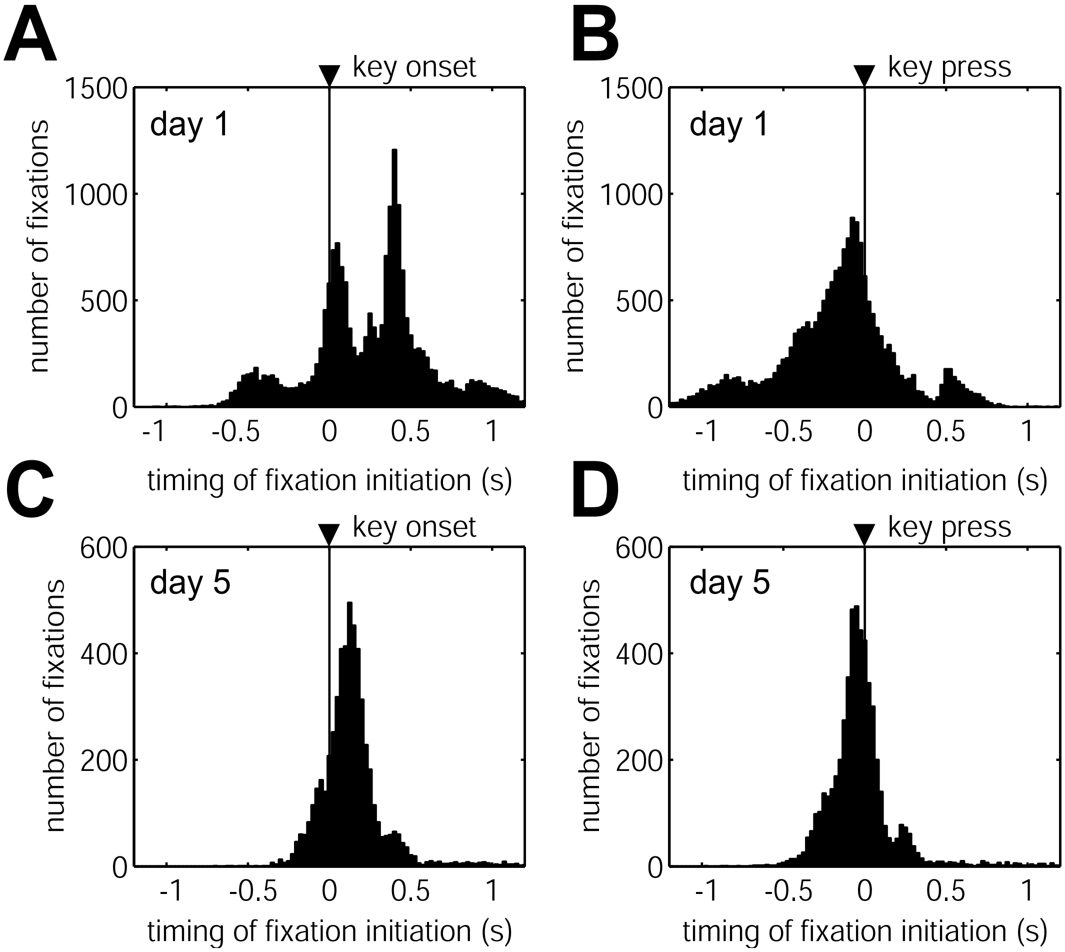

Figure 5.8. Subjects fixate targets prior to key press. Histograms of timing of fixation initiation for fixations ($N_{kpf} = 0, 1$, all success trials; 18,418 fixations on day 1, 5,412 fixations on day 5; bin size: 25 ms). (A) Timing on day 1 , with respect to key onset. The two higher peaks are at 0.050 s and 0.40 s. (B) Timing on day 1, with respect to key press. The peak is at -0.075 s. (C) Timing on day 5, with respect to onset. The peak is at 0.125 s. (D) Timing on day 5, with respect to key press. The peak is at -0.050 s. The reduction in variance of timing on day 5 is associated with the reduction in the movement times on day 5 compared to day 1. While the mean timing of the fixation initiation with respect to the key onset changed between days 1 and 5, the peak timing compared to the key press was maintained between days 1 and 5.

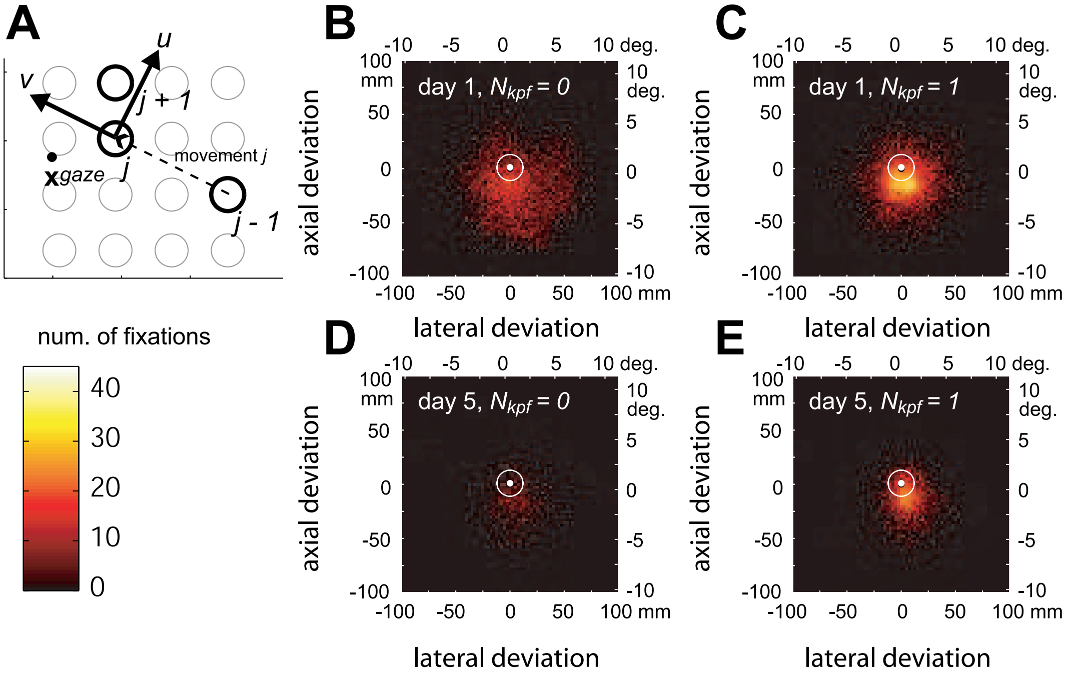

Figure 5.9. The spatial location of gaze fixation for fixations associated with a single target ($N = 0, 1$). (A) Generalized coordinate system $\{u, v\}$ for the histograms. Gaze location is referenced in comparison with the location of its associated target j. The coordinate system is then rotated so that v (" axial deviation") is aligned with movement j. The other, perpendicular axis u ("lateral deviation ") is positive in the direction of the next target ($j + 1$). (B-E) 2D histograms of target deviations in the generalized coordinate system (all success trials; 18,418 fixations on day 1, 5,412 fixations on day 5, bin size: 2.5 x 2.5 mm). Deviations are given in screen distances (mm, left ordinate, bottom abscissa) and corresponding gaze angle (deg., right ordinate, top abscissa). (B) Histogram for day 1, $N = 0$ (8,731 fixations): mean (u =5.8 mm, v=-16 mm), variance (σ_u^2= 730 mm^2, σ_v^2= 700 mm^2, σ_{uv}^2= -14.6 mm^2) . (C) Histogram for day 1, $N_{kpf} = 1$ (1,291 fixations): mean (5.9 mm, -16 mm), variance (297 mm^2, 356 mm^2, -11.4 mm^2) . (D) Histogram for day 5, $N_{kpf} = 0$ (9,445 fixations): mean (3.8 mm, -11 mm), variance (461 mm^2, 421 mm^2, -20.6 mm^2) . (E) Histogram for day 5, $N_{kpf} = 1$ (4121 fixations): mean (5.8 mm, -13 mm), variance (201 mm^2, 285 mm^2, -13 mm^2). Fixations tended to be located slightly short of the current target to be pressed (-v) and slightly towards the following target (+u).

Figure 5.9 shows histograms of the deviation of the gaze from the target for fixations ($N_{kpf} = 0, 1$), in a relative coordinate system of lateral (u) and axial (v) deviations with respect to the corresponding movement. On both days, the peaks of the distributions are centered on the edge of the target short of the target center (mean(v) < 0, $p < 0.0001$, all 4 distributions), inclined laterally towards the next target (mean(u) > 0, $p < 0.0001$, all 4 distributions). Compared to $N_{kpf} = 0$, the distributions are sharper for $N_{kpf} = 1$, which, because of their longer durations, may be more significant for visual feedback. The distributions were also sharper on day 5, suggesting that subjects learn to fixate this position.

For fixations $N_{kpf} \geq 2$, we tested three hypotheses for possible fixation targets: 1) center of (target) mass (COM), 2) closest target (closest target center to the fixation location) and 3) first target (of the target keys spanned by the fixation), see Section 5.1. Figures 5.10A-C show three examples for $N_{kpf} = 2$, $N_{kpf} = 3$ and $N_{kpf} \geq 4$, respectively. In Figure 5.10A and Figure 5.10B, fixations are close to COM. It becomes clearer in Figure 5.10C, where targets and COM are spatially separated, and fixations are closer to COM. These examples also reflect the statistics. Figure 5.10D shows the radial deviation from the candidate hypotheses. In order to test whether the differences were significant, an ANOVA with a fixed effect of fixation target and random effect of subjects was performed. If a significant main effect of fixation target was found, then a Tukey's HSD post-hoc test was performed to examine significant differences between the three fixation hypotheses. For all fixations $N_{kpf} \geq 2$, COM is as good candidate as closest target. The main fixed effect was significant ($F_{2,14} = 14.8$, $p < 0.001$). The post hoc test indicated that COM and closest target were significantly different from first target ($p < 0.001$) and ($p < 0.01$) respectively. There was no significant difference between COM and closest target (p=0.61). In Figure 5.10E, we only consider the subset of fixations $N_{kpf} > 2$ and where COM is separated from targets by at least 39 mm. The main fixed effect was significant ($F_{2,14} = 7.2$, $p < 0.01$). Here, COM is a better predictor than closest target ($p < 0.05$) and first target ($p < 0.01$). There was no significant difference between first target and closest target (p =0.62).

In summary, subjects mostly fixated single targets just before they were pressed. Such gaze was not fixated at target centers, but on the target edge,

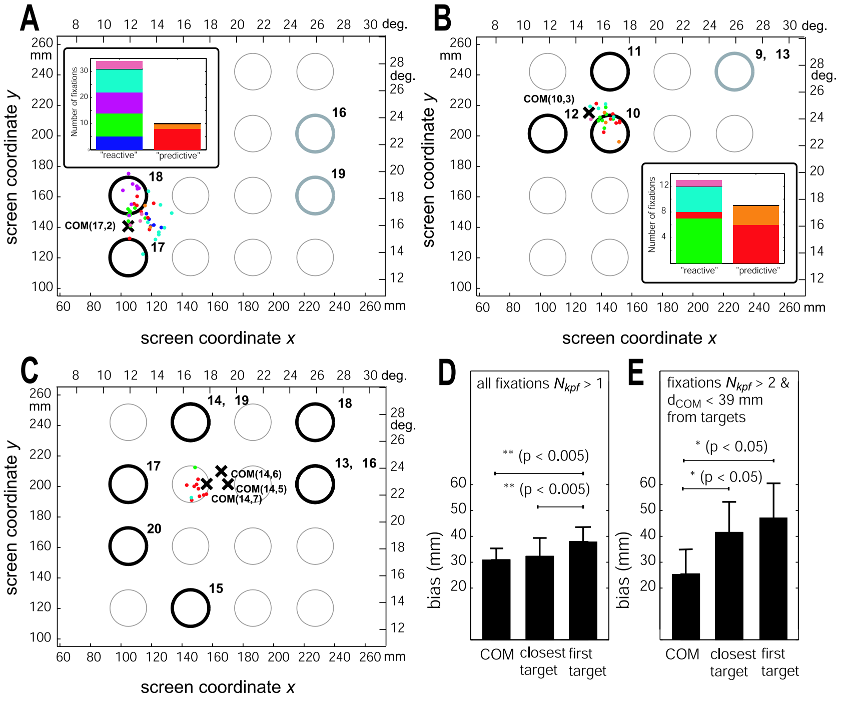

Figure 5.10. Fixations for multiple target presses are located at the center of mass (COM) of all of the pressed targets. Location of fixations $N_{kpf} > 1$ compared to target center-of-mass (COM(j, n), where j is the first rank order target, and n the total number of targets in the mass). (A) Example for $N_{kpf} = 2$, all subjects, last 9 trials of day 5. Dots indicate fixation locations (by subject colour) and the x marks COM(17, 2), sequence I (Targets included in COM are highlighted by bold circles). The stacked bars shows the number of fixations between subjects, divided between "reactive" (fixated during key presses 17-18) and "predictive" (fixated during key presses 16-17). (B) An example with $N_{kpf} = 3$, COM(10, 3), sequence I, where "reactive" was fixated during 10-12, and "predictive" during 9-11. (C) Examples of fixations with $N_{kpf} > 4$, all subjects, all success trials on all days, sequence II. Fixations were initiated when targets 13 or 14 was active. Locations of COM(14, 5), COM(14, 6) and COM(14, 7) are marked. (D) Average bias (radial deviation) from three candidate fixation targets: COM, closest target and first target, for all fixations $N_{kpf} > 1$. (E) The same as (D), but for the subset $N_{kpf} > 2$ and where the distance between COM and the closest target (dCOM) was larger than the shortest movement (39 mm). In (D-E), significant differences between the fixation location hypotheses were determined using a Tukey's post-hoc test after a significant main effect with an ANOVA.

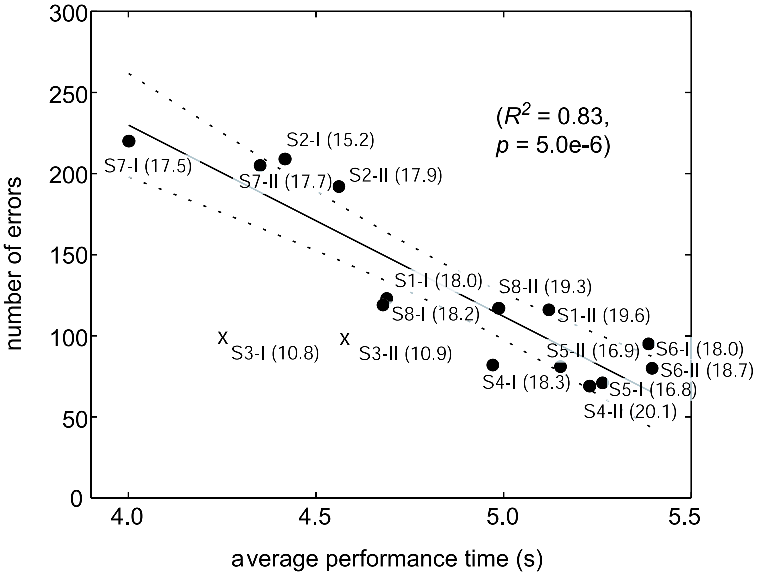

Figure 5.11. Speed accuracy trade-off. The mean performance times of the 25 fastest movements (when subjects are presumably most motivated) on day 4 and 5 (when subjects are most proficient) are plotted versus the total number of misses on day 4 and 5. Each data point is for a single subject and sequence (labeled SX-I and SX-II). The number in parentheses indicates the mean number of gaze shifts of the 25 trials. The fitted line (dotted lines: 95% confidence interval) is based on all subjects (marked as dots) with the exception of one, classified as an outlier (marked by x). This outlier had the lowest mean number of gaze shifts for both sequence I and II.

with an inclination towards the next target. This effect was more pronounced in late training. In late training, it was also more common that fixations spanned several key presses. The fixation location was towards the center-of-mass of the pressed target keys rather than to one of the keys to be pressed.

5.2.3 Gaze and performance

Figure 5.11 indicates the existence of a speed-accuracy trade-off. Average performance time (of the 25 fastest success trials on day 4 and 5) versus number of errors is shown for the last two days of training for each sequence. The correlation coefficient R^2 was 0.83 ($p < 0.0001$). Two subjects (S2 and S7) performed significantly faster than the others at the cost of a much higher error rate. Another subject (S3), classified as an outlier (marked by x), violated the trade-off by being both accurate and fast. This subject was also using significantly less gaze shifts on average than others; 10.8 (S3) versus (17.4 ± 1.1 SD) (others) for sequence I, and 10.9 (S3) versus (18.5 ± 1.2 SD) others for sequence II.

As reported in the section on task performance, accuracy is context dependent. We tested if the population RMFs could be predicted by our Bayesian model of gaze-dependent certainty (dynamic updating of spatial representation, see Section 5.1). The model assumes that the estimate of target locations is subject to diffusion, caused by drift. The estimate at the previous time step is taken as the prior, which is multiplied with the likelihood of the visual input to obtain the posterior. The certainty of this likelihood depends on the eccentricity of the target. We predict that the measured accuracy of hand movements (RMF) should roughly correlate with the computed certainties of target estimates, at the time of target press. For our predictions, we tested three assumptions of gaze trajectories: 1) "target", assuming gaze fixations of each target, 2) "mode", assuming the mode location of measured fixations, and 3) "random", a control assumption using random locations.

Figure 5.12A shows the dependence of RMF on σ_j^{-2} (the certainty of a target location i, with rank order j at key press time $t = t_j$) computed by our standard assumption ("mode"). The correlation ($R^2 = 0.37$, $p < 0.0001$) was done for 37/40 data points (excluded points were σ_1^{-2} for both sequences, as they are sensitive to initial conditions of both hand movement and gaze, and σ_{11}^{-2} of sequence II, as the corresponding RMF has a variance significantly larger than any other data point (see Figure 5.3). Figure 5.12B shows that this prediction is robust for both our assumptions "target" and "mode", also when assumed trajectories are perturbed, but does better than the control assumption, any arbitrary trajectory. That means that our model can roughly predict the accuracy of a target given a general gaze trajectory, but fails to be sensitive for variations in gaze and RMF (for more accurate estimation of $p(\text{miss} \mid \text{gaze trajectory})$, gaze needs to be controlled).

In summary, we found a trade-off between speed and accuracy. Notably, a

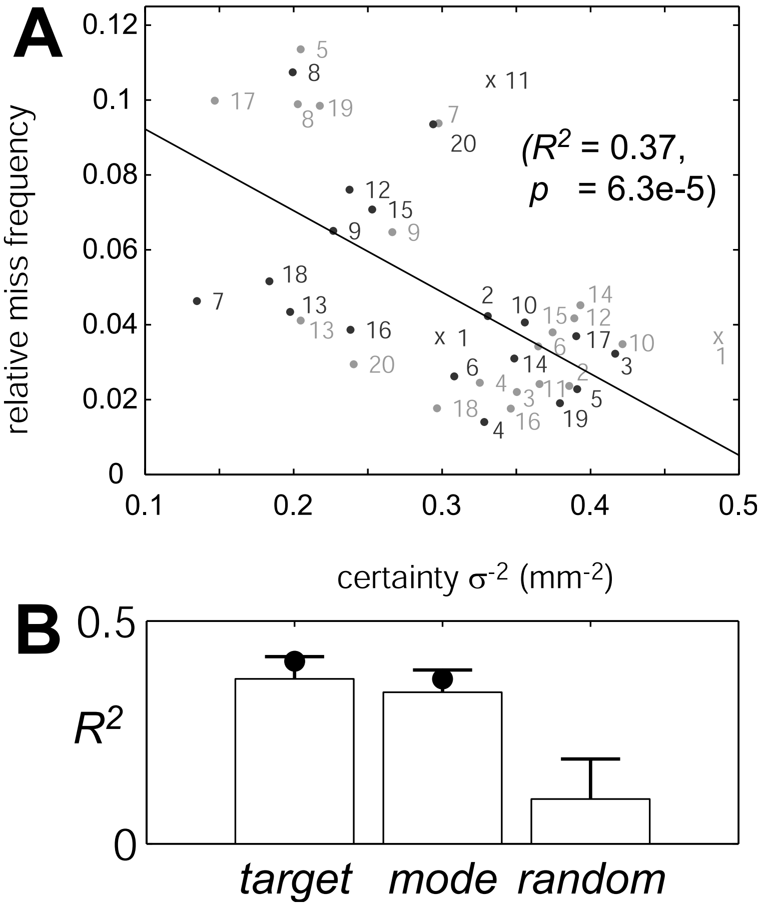

Figure 5.12. The Bayesian model of dynamic updating of spatial location is able to predict the relative miss frequency of targets. (A) Scatter plot of model prediction (certainties σ_j^{-2}), using the standard parameters of the model (see text) versus relative miss frequencies (RMF). Light and dark gray indicates sequence I and II respectively, numbers are rank orders j . Data points marked with x have been removed from the data fit (see text). (B) Robustness of model prediction in terms of R^2, for different assumptions of gaze trajectory: " target " (fixating each target center), " mode " (mode of measured gaze, standard assumption) and a control assumption of "random" (a sequence of 20 fixations with random locations within panel). The bullets indicate results for the assumed gaze trajectories, while the bars with standard deviations show results for the perturbed assumptions, (randomly by an SD of 10 mm in both horizontal and vertical (100 times)).

single subject using significantly less number of gaze shifts could perform significantly more accurately at higher speeds than other subjects. Also, we found that our Bayesian model of dynamic updating of spatial representation could predict the general pattern of relative miss frequencies of subjects.

5.3. Discussion

All subjects improved their performance in line with what was asked of them - to move as fast as possible. Even less well-performing (i.e. slower) subjects sped up performance considerably. As they were subject to a repetitive task, proceduralization made it possible to execute without much hesitation or cognitive effort on later days of training. The cost of poor accuracy was low in the present task. The only cost was the indirect consequence of missing a target - the time lost from the miss until the subject reacted and could resume. Accordingly, we did not observe a decrease in errors with experience. On the contrary, the error rate increased slightly. Accuracy acquired from experience was presumably traded for speed.

5.3.1 Gaze shifts

Earlier studies have shown that targets are fixated in preparation of manipulation [6, 138, 76] . Our task confirms these results with some modification. When execution was slow enough, gaze responded to each target. However, as the number of key presses per time increased, the change in gaze frequency was only modest, implying some constraints for gaze shifts. Subjects modified gaze location rather than frequency and timing. It seems that subjects have to make a trade-off between loss of accuracy due to omission of visual input during gaze shifts on one hand, and loss of accuracy due to target eccentricity on the other. Peripheral vision enables subjects to update locations of multiple targets. This is consistent with our observations that 1) fixations associated with a single target were also inclined towards the next target, and 2) there were a significant fraction of fixations where several manipulated targets were observed from their center-of-mass. It is known earlier that subjects fixate center-of-mass when several target stimuli appear simultaneously on a screen [45, 131]. In our experiment, however, the fixation of center-of-mass requires a real-time, context-sensitive decision. It has be to predicted for a subsequence of targets presented one-by-one, and is thus a more complex decision than for presentation of multiple, simultaneous targets.

The cost of gaze shifts is presumably functional, rather than metabolic, since energy loss is much smaller compared to that of the hand movements. The loss of visual feedback during saccades may be one factor: for example, if a subject fixates 20 targets in 4 seconds, 1 second is lost by saccades of 50 ms duration. Saccades are also inherently inaccurate due to noise in ocular motor commands, multiplying with amplitude [59] , introducing uncertainty in the short term.

It is also of note that subjects, even after 5 days of training, make responsive gaze shifts at all, instead of total reliance on somatosensation and procedural memory. In this task, there is no alternative use of vision. (A piano player, for instance, needs vision for secondary tasks like reading notes). It is possible that the response of gaze is habitual. However, the adaptive change of gaze with increasing speed implies a function of gaze also in late learning; to enhance accuracy, which can be traded for speed.

5.3.2 Dynamic updating of spatial representation

We have proposed a Bayesian model of dynamic updating of spatial representation [24] , which could predict context dependent accuracy of hand movements. The model is based on two main assumptions: 1) that the maintained spatial map degrades over time by drift, and 2) that the spatial map is reinforced by gaze, depending on eccentricity. We speculate whether or not the impact of gaze on manual performance could be understood by this model. It is known from neurophysiology that target locations are held in memory by several brain areas, also after new gaze shifts [166, 24, 167, 158] . The questions are 1) at what speed this short term memory degrades, and 2) how the remapping mechanisms following gaze shifts would affect this memory. Future experiments explicitly controlling timing and fixations could determine the exact relationship between gaze trajectory and manual accuracy. It is likely that several feedback modalities are utilized for spatial estimations [24], and not vision exclusively. Our model is passive in the sense that predictive components play no part in computing the estimates. However, our model may be more reflective on the fine tuning of accuracy, for which visual feedback is particularly useful. This fine-tuning would also account for the fact that vision respond also in late learning: the maximum accuracy of the learned non-visual, proprioceptive and feedforward-driven of the arm is likely to be bounded lower than for vision.

Referencing between locations of the index finger and target may also important, rather than just only the absolute positions of targets.

5.3.3 Gaze strategies

The Bayesian model of dynamic updating of spatial representation points to an optimal gaze strategy to maximize accuracy. Recently, similar models have successfully predicted optimal behaviour in visual search tasks [122, 139] . The tendency of fixating center-of-mass of targets appearing rapidly is predicted by our model, then the cost for refixation is greater than that of eccentricity of targets. Seen as an objective function for maximizing manual accuracy, it remains to be seen whether subjects learn a globally optimal strategy, or gets stuck in a global optimum. Since performance is dependent on gaze history, the objective function has many local optima, so finding a global optimum may be hard. It is also possible that several local optima are approximately equal in value. A subject of future study is to study manual accuracy in experiments where gaze is controlled, for example for a stationary gaze.

An alternative possibility is that optimality is driven on a more global level, learning decisions and coordination of eye and arm movements simultaneously.

5.4. Conclusion

We have in this study for the first time demonstrated that with extensive motor learning humans do not simply change arm kinematics, but also change gaze strategy. As motor performance sped up, the number of gaze shifts became fewer, reducing to below the number of targets to which the hand moved. The location of gaze fixations also shifted, tending towards the center-of-mass of multiple targets. This observation suggest that subjects need to make a trade-off between loss of vision due to gaze shifts on one hand, and due to target eccentricity on the other. Future experiments designed to evaluate theories of optimal gaze strategies will learn us more how visual feedback is used for learning and control of movements. An accurate model of gaze-dependent accuracy may also elucidate neural mechanisms of areas involved in spatial estimation.

Furthermore, most experimental tasks of eye-hand coordination report on a coupling between eye and hand. Here, we see a gradual decoupling between sequences of eye and hand responses. This also implies that two sequences need to be stored in memory, rather than one.

Overall, this study points out that vision also plays a role also in controlling overtrained manual motor skills, which implies learning gaze strategies of non-trivial optimality.

Appendix

Here, we verify the sensitivity of the model to parameter variation, and justify the parameter set in terms of known neurophysiology and psychology.

5.A. Robustness to parameter variation

Figure 5.13 shows sensitivity of the prediction (Figure 5.12) with respect to model parameters, in terms of R^2. (Note that acuity parameter A_0 only change the scale of certainties, and does not affect our prediction, since RMF is relative). For gaze to be important for hand accuracy, the speed of change of accuracy (diffusion rate ν) must scale with the time scale of the task. Thus, when diffusion is too slow, gaze shifts do not matter, producing weak correlations for low ν. The same principle applies for scale parameter of distal acuity σ^{gaze} - if eccentric (peripheral) targets can be estimated as good as central (foveal), gaze shifts are not necessary. On the other hand, if σ^{gaze} is small, peripheral targets become unimportant and subjects would have no choice but to fixate every target. Our model has tolerance for a variable feedback delay τ^{delay} up to 200 ms, which makes our assumption about recall timing of target estimate (see methods) less important. Also, saccade duration can be varied within reasonable bounds without unduly affecting the prediction.

5.B. Constraints from known neurophysiology and psychology

A study of monkey saccades suggests that a spatial representation can be acquired and maintained in the frontal eye field for several minutes without update [167] . Our model also assumes a prior estimate, which represent the memory acquired map. The dynamic updating is not crucial, it only enhances the quality of the map. In a sequential planning task, [158] showed that accuracy of target pointing increased with presentation time in the presence of distractors, continuously over up to 8 seconds of presentation. In our model, the diffusion rate reported is fast enough to make spatial accuracy vary in real time, but slow enough to make gaze history important.

Early studies have shown a linear relationship between eccentricity and visual resolution [173, 144, 8] , i.e. an inverse linear relationship between accuracy and eccentricity. Some studies have also shown degraded performance of pointing with respect to target eccentricity, stressing the difference of foveal and peripheral vision, rather than a continuous correlation [1, 158] . For convenience, we used a Gaussian function, where narrow and broad peaks points to non-linear and linear relationships, respectively. Our prediction was optimal for a peak width of 6 deg. (target matrix is 16 deg. wide), implying a degree of non-linear relationship. A narrower peak increases the model's sensitivity to variability of gaze location, which cannot be predicted without a more accurate measure of hand accuracy, which may be the reason of the steep loss of correlation for narrower peaks in our study.

The latency from perceived visual stimulus to implementation of spatial estimate in computation of hand movement depends on delays in the neural pathways, and at what time the estimate is recalled. We assumed a recall for immediate, online correction. The other possibility is a recall during one or several stages of movement planning, for example during the initiation of the hand movement

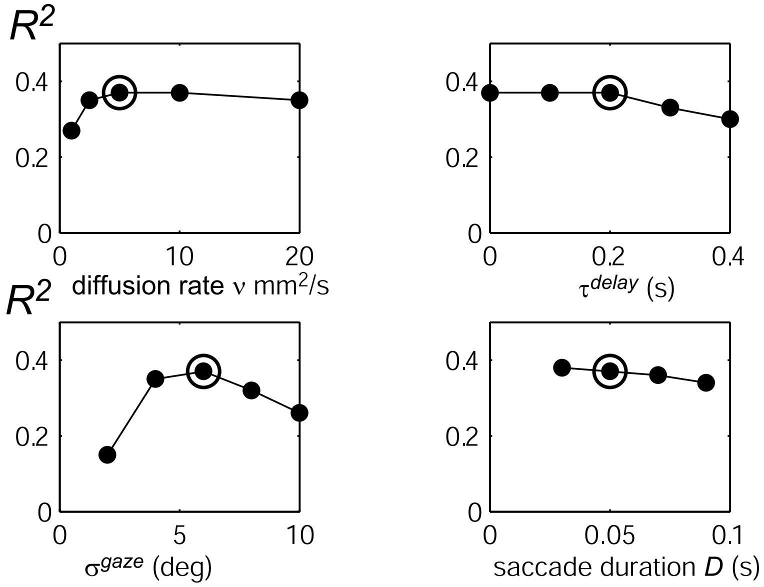

Figure 5.13. Prediction robustness in terms of R^2 to parameters: diffusion rate, feedback delay, gaze width and saccade duration. The encircled points are the standard assumption ("mode") as in Figure 5.12.

(previous key press). As the model prediction was robust for 100's of milliseconds in delay, the model is forgiving to this assumption, to some degree.

Chapter 6

Conclusions

In this thesis, we explored two different lines of work regarding real-time constraints of human motor control and learning. In the first, a computational study, we investigated the real-time interactions of modalities with different latencies under a modular, actor-critic framework, in the learning and control of reaching movements. Population coded outputs were combined with a softmax function. After learning, performance depended on the faster modality alone. Further, we found that a preacquired, slower visual module display different roles during training of a faster somatosensory expert modality: in early learning, the visual module acts as a guide for the somatosensory module. Late in learning, the somatosensory module predominates to outperform the visual module, but the visual module flexibly takes over when the somatosensory module goes out of control.

In the second, an experimental study, we examined the long term changes of gaze behaviour in a manual, sequential task. We found that as subjects performed faster, gaze behaviour shifted from fixating each target to shifting key locations, commonly the center-of-mass of a subsequence of targets. We also found that manual accuracy was context dependent, and could be explained by a gaze-dependent Bayesian model of dynamic updating of spatial representations.

This work contribute to developing a full theory of how cortical visual and somatosensory feedback loops are utilized for learning and control of goal driven, complex motor behaviour - indeed a very complex and challenging problem. As argued throughout this thesis, studying the temporal constraints, and timing of behavioural events and neural firings could provide important clues for understanding such a theory. Below we point to some future work that are steps in this direction, and conclude with a summary of the contributions of this thesis.

6.1. Future directions

Our computational study of real-time, multi-modal combination addresses several important topics to be investigated in neuroscience:

1) Physiology of the multi-modular loop BG-TC system. More electrophysiological studies are needed with cell recording of several areas in the basal ganglia-thalamocortical system (BG-TC) simultaneously [57], for different stages of motor skill learning. Together with imaging experiments, such studies would teach us, for example about how different loop circuits communicate, or whether the actor-critic learning framework, as argued in this thesis, is applicable to motor learning as well.

2) Combination of modalities under reinforcement learning. This is a hard theoretical problem, and the neural mechanisms are largely unknown. The anatomical loop structure [2] suggests that, within a reinforcement learning framework, the output of loops would be evaluated separately. Separate evaluation was also successful in our previous, discrete model [123], but here, we found it necessary to relax this constraint, and assume a combined evaluation of outputs. For separate evaluation to work, each loop must receive sufficient, timely information about the state of other loops. For example, the pre-supplementary area or the cingulate motor area may realize this function between prefrontal and motor loops [123].

3) Feedback versus feedforward modules. For the sake of simplicity, feedforward was left out in our study. However, feedforward modules could easily be included in our model, and the relative utility of feedback versus and feedforward control could be studied. Using a more realistic, muscle model plant with accurate noise models [22] based on experimental findings [77, 130] could be used to study the interaction of slow/clean versus fast/noisy control modules in speed-accuracy trade-off.

4) Comparison to state estimation models. The recently popular optimal feedback control paradigm [163, 161] may also explain the larger gain of faster modules. An important field of work is to quantify the limitations of feedback control models, and test if alternative control paradigms are competitive. For example, recursive state estimation may be costly in real-time. Also, Kalman filter-based frameworks assume available system and sensor models, and are limited to linear systems.

Further, our finding of decoupling of gaze and hand in late learning of manual sequential skills, make us ask several questions:

1) Cost of gaze shifts. The observation that subjects abandons the strategy of fixating each manual target in fast, mature execution suggests that there is a limit to how fast gaze shifts can be processed and integrated. The visual input of the new gaze location must be remapped and recalibrated with respect to the previous gaze location [41]. Visual feedback is also lost during the actual saccade [19]. Faced with these constraints, subjects are left with the option to rely on peripheral vision. Experiments with controlled gaze behaviour would improve our understanding of how gaze-dependent vision affects manual accuracy, and to what lengths subjects adapt to improve feedback.

2) Neural mechanisms of dynamic updating of spatial representation [24] . An active area in neurophysiology is to understand how space is represented, particularly in the intraparietal areas [24]. A precise understanding of gaze behaviour, as our proposed Bayesian model may help to understand real-time patterns of neural activity.

3) Multi-modal sequence learning. The decoupling of visual and motor sequences implies that two different sequences need to be stored in memory. How gaze sequences are chunked, and how that affect sequential memory would elucidate our understanding of procedural memory. Of note are also the parallel paths between the prefrontal cortex and a) the frontal eye field through the supplementary eye field (SEF) [72, 71], and b) the primary motor cortex through the supplementary motor area (SMA). The pre-SEF and the pre-SMA, respectively, may store differential sequences, coordinated by the prefrontal cortex.

6.2. Summary of contributions

Combining modalities with different latencies Our main contributions of this study are as follows:

1) **A successful real-time implementation of a basal ganglia motor learning model**. In contrast to most actor-critic models of the basal ganglia, which typically models event-driven conditioning tasks, we here demonstrated successful learning of real-time, joint torque controlled arm movements with delayed feedback, which may fuel the discussion of the function of the basal ganglia in motor control.

2) **Population coded combination of modules with different latencies by a softmax function**. Using population-coded outputs for each modality, combined by a softmax function, was demonstrated to be a flexible way to do action selection in motor control. Instead of using explicit gating, proper input-output associations are strengthened by reinforcement learning. The softmax function further enhances the influence of the more sharply distributed population code (often the faster modality), while the broader population code (often the slower modality) is suppressed. The combination should be equally useful for other differences in feedback qualities, like signal-to-noise ratios.

3) **Robust modular switching in visuomotor skills**. Our framework was shown to be flexible for exploiting the most useful controller among redundant feedback. In our example, a slow, but robust, general-purpose controller ("the visual module") served as a guide to speed up learning of a fast specialized controller ("the somatosensory module", who otherwise would have to learn under "motor babbling", for which learning is painfully slow in high-dimensional systems). Once trained, the somatosensory module would overtake control and outperform the visual module. However, because the somatosensory module only is effective in its expert regime, the visual module would resume control whenever the arm went out of control. Thus, the visual controller plays the roles of tutor (early training) and safe-guard (late-training).

4) **The importance of feedback delays in memory transfer**. Our results suggest that the suboptimality of long delayed feedback (e.g. vision) may provide the indirect effect of memory transfer to modalities of shorter delay (e.g. somatosensation). In our framework, the output of faster feedback loops eventually overrides that of slower feedback with training. In a brain, this would enable the agent to transfer control from slower, conscious and working memory-driven control loops to faster, unconscious procedural memory, so that the conscious mind can be pre-occupied with other things.

5) **A quantitative picture of the effect of cortical feedback delays on motor performance**. Our results across 0-200 ms of feedback delay space provide a quantitative picture about the limitation of direct, delayed feedback control. Consistent with behavioural data on motor tasks [112, 92], the performance starts to degrade significantly from 50 ms.

6) **Temporal discounting in reinforcement learning in the context of motor skills**. Temporal discounting of reward makes early delivery worth more than late delivery of the same amount of reward. In the context of motor control, this effect may explain why motor skills are most often associated with speedup of performance, although such speedup is not an explicit learning goal [5, 64].

Gaze strategies in sequential hand movements Our main contributions of this study are as follows:

1) **A change in gaze behaviour in long-term learning of sequential hand movement**. The experimental results confirmed our hypothesis that the manual, behavioural change in long-term skill learning should bring about a change in gaze strategy. Somewhat contrary to intuition, as performance speeds up, the hand must wait for the eye, as a very frequently refixating eye cannot process feedback in time. In early learning, subjects fixated almost every target, but late in learning, subjects refixated less number of times than there were targets.

2) **Predictive fixations during fast, manual execution**. Late in learning, subjects fixated predictively in two ways. In some cases, fixations of targets

were inclined towards the next target. In other cases, subjects would fixate the center-of-mass of a subsequence of targets. This suggests that subjects need to rely more on peripheral visual feedback during fast manual execution, rather than refixation.

3) **Context-dependent manual accuracy in sequential hand movement**. We report that the accuracy of aimed arm movements embedded in the pressing task were context-dependent. The accuracy of rank order-specific targets were consistent across subjects and training.

4) **A gaze-dependent, Bayesian model of dynamic updating of spatial representation**. We proposed a Bayesian model of dynamic updating of spatial representations. The model provides a method to estimate the real-time dependent accuracy of visual targets, and could be used to predict behaviour and possibly neural firing patterns. It can also be used to test the relative importance of foveal and off-foveal vision. Our model is supported as it could to some degree predict the context dependent accuracy of hand movements in the present study.

Bibliography

[1] R. A. Abrams, D. E. Meyer, and S. Kornblum. Eye-hand coordination: oculomotor control in rapid aimed limb movements. *Journal of Experimental Psychology: Human Perception and Performance*, 16(2):248–67, 1990.

[2] G. E. Alexander and M. D. Crutcher. Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends in Neurosciences*, 13(7):266–71, 1990.

[3] S. Amari. Natural gradient works efficiently in learning. *Unsupervised Learning: Foundations of Neural Computation*, 1999.

[4] R. A. Andersen, L. H. Snyder, D. C. Bradley, and J. Xing. Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience*, 20(1):303–330, 1997.

[5] J.R. Anderson. *Learning and memory.* John Wiley & Sons, Singapore, 1995.

[6] R. W. Angel, W. Alston, and H. Garland. Functional relations between the manual and oculomotor control systems. *Experimental Neurology*, 27(2):248–57, 1970.

[7] S. Anstis. Picturing peripheral acuity. *Perception*, 27(7):817–825, 1998.

[8] S.M. Anstis. A chart demonstrating variations in acuity with retinal position. *Vision Research*, 14(7):589–92, 1974.

[9] I. Bar-Gad, G. Havazelet-Heimer, J. A. Goldberg, E. Ruppin, and H. Bergman. Reinforcement-driven dimensionality reduction - a model for

information processing in the basal ganglia. *Journal of Basic Clinical Physiology and Pharmacology*, 11(4):305–320, 2000.

[10] A. Barto. Adaptive critics and the basal ganglia. In J. Houk, J. Davis, and D. Beiser, editors, *Models of Information Processing in the Basal Ganglia*, pages 215–232. MIT Press, Cambridge, MA., 1995.

[11] J. F. Bates and P. S. Goldman-Rakic. Prefrontal connections of medial motor areas in the rhesus monkey. *The Journal of Comparative Neurology*, 336(2):211–228, 1993.

[12] R. E. Bellman. *Dynamic Programming*. Courier Dover Publications, 2003.

[13] R. A. Berman, L. M. Heiser, C. A. Dunn, R. C. Saunders, and C. L. Colby. Dynamic circuitry for updating spatial representations. iii. from neurons to behavior. *Journal of Neurophysiology*, 98(1):105, 2007.

[14] G. S. Berns and T. J. Sejnowski. A computational model of how the basal ganglia produce sequences. *Journal of Cognitive Neuroscience*, 10(1):108–21, 1998.

[15] D. P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 1995.

[16] D. P. Bertsekas and J. N. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, 1996.

[17] B. Biguer, C. Prablanc, and M. Jeannerod. The contribution of coordinated eye and head movements in hand pointing accuracy. *Experimental Brain Research*, 55(3):462–9, 1984.

[18] F. Bissmarck, H. Nakahara, K. Doya, and O. Hikosaka. Responding to modalities with different latencies. In K. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems*, volume 17, pages 169–176, Vancouver, Canada, 2005. Camebridge, MA: MIT Press.

[19] B. Bridgeman, D. Hendry, and L. Stark. Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, 15(6):719–22, 1975.

[20] C. J. Bruce and M. E. Goldberg. Primate frontal eye fields. i. single neurons discharging before saccades. *Journal of Neurophysiology*, 53(3):603–635, 1985.

[21] E. Burdet, R. Osu, D. W. Franklin, T. E. Milner, and M. Kawato. The central nervous system stabilizes unstable dynamics by learning optimal impedance. *Nature*, 414(6862):446–9, 2001.

[22] E. Burdet, K. P. Tee, I. Mareels, T. E. Milner, C. M. Chew, D. W. Franklin, R. Osu, and M. Kawato. Stability and motor adaptation in human arm movements. *Biological Cybernetics*, 94(1):20–32, 2006.

[23] L. G. Carlton. Processing visual feedback information for movement control. *Journal of Experimental Psychology: Human Perception and Performance*, 7(5):1019–30, 1981.

[24] C. L. Colby and M. E. Goldberg. Space and attention in parietal cortex. *Annual Review of Neuroscience*, 22:319–49, 1999.

[25] M. Connolly and D. Van Essen. The representation of the visual field in parvicellular and magnocellular layers of the lateral geniculate nucleus in the macaque monkey. *The Journal of Comparative Neurology*, 226(4):544–564, 1984.

[26] A. Cowey and E. T. Rolls. Human cortical magnification factor and its relation to visual acuity. *Experimental Brain Research*, 21(5):447–454, 1974.

[27] J. D. Crawford, W. P. Medendorp, and J. J. Marotta. Spatial transformations for eye-hand coordination. *Journal of Neurophysiology*, 92(1):10–9, 2004.

[28] C. A. Curcio and K. A. Allen. Topography of ganglion cells in human retina. *The Journal of Comparative Neurology*, 300(1):5–25, 1990.

[29] C. A. Curcio, K. R. Sloan, O. Packer, A. E. Hendrickson, and R. E. Kalina. Distribution of cones in human and monkey retina: individual variability and radial asymmetry. *Science*, 236(4801):579–582, 1987.

[30] P. M. Daniel and D. Whitteridge. The representation of the visual field on the cerebral cortex in monkeys. *The Journal of Physiology*, 159(2):203, 1961.

[31] N. Daw. *Reinforcement learning models of the dopamine system and their behavioural implications*. PhD thesis, Carnegie Mellon University, 2003.

[32] N. D. Daw, S. Kakade, and P. Dayan. Opponent interactions between serotonin and dopamine. *Neural Networks*, 15(4-6):603–616, 2002.

[33] N. D. Daw and D. S. Touretzky. Long-term reward prediction in td models of the dopamine system. *Neural Computation*, 14(11):2567–83, 2002.

[34] S. Deneve, J. R. Duhamel, and A. Pouget. Optimal sensorimotor integration in recurrent cortical networks: A neural implementation of kalman filters. *Journal of Neuroscience*, 27(21):5744, 2007.

[35] B. M. Dow, A. Z. Snyder, R. G. Vautin, and R. Bauer. Magnification factor and receptive field size in foveal striate cortex of the monkey. *Experimental Brain Research*, 44(2):213–228, 1981.

[36] K. Doya. What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks*, 12(7-8):961–974, 1999.

[37] K. Doya. Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology*, 10(6):732–9, 2000.

[38] K. Doya. Reinforcement learning in continuous time and space. *Neural Computation*, 12(1):219–45, 2000.

[39] J. Doyon and H. Benali. Reorganization and plasticity in the adult brain during learning of motor skills. *Current Opinion in Neurobiology*, 15(2):161–7, 2005.

[40] J. Doyon, A. M. Owen, M. Petrides, V. Sziklas, and A. C. Evans. Functional anatomy of visuomotor skill learning in human subjects examined with positron emission tomography. *European Journal of Neuroscience*, 8(4):637–48, 1996.

[41] J. R. Duhamel, C. L. Colby, and M. E. Goldberg. The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, 255(5040):90–92, 1992.

[42] R. O. Duncan and G. M. Boynton. Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*, 38(4):659–671, 2003.

[43] S. A. Engel. Retinotopic organization in human visual cortex and the spatial precision of functional mri. *Cerebral Cortex*, 7(2):181–192, 1997.

[44] C. B. Ferster and B. F. Skinner. *Schedules of reinforcement.* Appleton-Century-Crofts New York, 1957.

[45] J. M. Findlay. Global visual processing for saccadic eye movements. *Vision Research*, 22(8):1033–45, 1982.

[46] J. D. Fisk and M. A. Goodale. The organization of eye and limb movements during unrestricted reaching to targets in contralateral and ipsilateral visual space. *Experimental Brain Research*, 60(1):159–78, 1985.

[47] P.M. Fitts. Perceptual-motor skill learning. In AW Melton, editor, *Categories of human learning*, pages 243–285. Academic Press, New York, 1964.

[48] J. R. Flanagan and R. S. Johansson. Action plans used in action observation. *Nature*, 424(6950):769–771, 2003.

[49] T. Flash and N. Hogan. The coordination of arm movements: an experimentally confirmed mathematical model. *Journal of Neuroscience*, 5(7):1688, 1985.

[50] A. Floyer-Lea and P. M. Matthews. Distinguishable brain activation networks for short- and long-term motor skill learning. *Journal of Neurophysiology*, 94(1):512–8, 2005.

[51] A. J. Foulkes and R. C. Miall. Adaptation to visual feedback delays in a human manual tracking task. *Experimental Brain Research*, 131(1):101–10, 2000.

[52] S. Funahashi, C. J. Bruce, and P. S. Goldman-Rakic. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61(2):331–349, 1989.

[53] H. Gomi, M. Shidara, A. Takemura, Y. Inoue, K. Kawano, and M. Kawato. Temporal firing patterns of purkinje cells in the cerebellar ventral paraflocculus during ocular following responses in monkeys i. simple spikes. *Journal of Neurophysiology*, 80(2):818–31, 1998.

[54] E. Gould, N. J. Woolf, and L. L. Butcher. Cholinergic projections to the substantia nigra from the pedunculopontine and laterodorsal tegmental nuclei. *Neuroscience*, 28(3):611–23, 1989.

[55] S. T. Grafton, E. Hazeltine, and R. B. Ivry. Abstract and effector-specific representations of motor sequences identified with pet. *Journal of Neuroscience*, 18(22):9420, 1998.

[56] A. M. Graybiel. The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory*, 70(1-2):119–36, 1998.

[57] A. M. Graybiel. The basal ganglia: learning new tricks and loving it. *Current Opinion in Neurobiology*, 15(6):638–44, 2005.

[58] S. N. Haber, K. S. Kim, P. Mailly, and R. Calzavara. Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. *Journal of Neuroscience*, 26(32):8368–76, 2006.

[59] C. M. Harris and D. M. Wolpert. Signal-dependent noise determines motor planning. *Nature*, 394(6695):780–4, 1998.

[60] M. Haruno and M. Kawato. Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fmri examination in stimulus-action-reward association learning. *Neural Networks*, 19(8):1242–1254, 2006.

[61] M. Haruno, D. M. Wolpert, and M. Kawato. Mosaic model for sensorimotor learning and control. *Neural Computation*, 13(10):2201–20, 2001.

[62] O. Hikosaka, H. Nakahara, M. K. Rand, K. Sakai, X. Lu, K. Nakamura, S. Miyachi, and K. Doya. Parallel neural networks for learning sequential procedures. *Trends in Neurosciences*, 22(10):464–71, 1999.

[63] O. Hikosaka, K. Nakamura, K. Sakai, and H. Nakahara. Central mechanisms of motor skill learning. *Current Opinion in Neurobiology*, 12(2):217–22, 2002.

[64] O. Hikosaka, M. K. Rand, S. Miyachi, and K. Miyashita. Learning of sequential movements in the monkey: process of learning and retention of memory. *Journal of Neurophysiology*, 74(4):1652–61, 1995.

[65] O. Hikosaka, M. K. Rand, K. Nakamura, S. Miyachi, K. Kitaguchi, K. Sakai, X. Lu, and Y. Shimo. Long-term retention of motor skill in macaque monkeys and humans. *Experimental Brain Research*, 147(4):494–504, 2002.

[66] M. Honda, M. P. Deiber, V. Ibanez, A. Pascual-Leone, P. Zhuang, and M. Hallett. Dynamic cortical involvement in implicit and explicit motor sequence learning. a pet study. *Brain*, 121(11):2159–2173, 1998.

[67] J. C. Houk, J. L. Davis, and D. G. Beiser. *Models of Information Processing in the Basal Ganglia*. MIT Press, 1995.

[68] J.C. Houk, J.L. Adams, and A.G. Barto. A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J.C. Houk, J.L. Davis, and D.G. Beiser, editors, *Models of Information Processing in the Basal Ganglia*, pages 249–270. MIT Press, Cambridge, MA, 1995.

[69] H. Imamizu, S. Miyauchi, T. Tamada, Y. Sasaki, R. Takino, B. Putz, T. Yoshioka, and M. Kawato. Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature*, 403(6766):192–5, 2000.

[70] A. W. Inhoff and J. Wang. Encoding of text, manual movement planning, and eye-hand coordination during copytyping. *Journal of Experimental Psychology: Human Perception and Performance*, 18(2):437–48, 1992.

[71] M. Isoda. Context-dependent stimulation effects on saccade initiation in the presupplementary motor area of the monkey. *Journal of Neurophysiology*, 93(5):3016–22, 2005.

[72] M. Isoda and J. Tanji. Participation of the primate presupplementary motor area in sequencing multiple saccades. *Journal of Neurophysiology*, 92(1):653–9, 2004.

[73] R. Jacobs, M. Jordan, and A. Barto. Task decomposition through competition in a modular connectionist architecture: The what and where in vision tasks. *Cognitive Science*, 15:219–250, 1991.

[74] D. Joel, Y. Niv, and E. Ruppin. Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Networks*, 15(4-6):535–47, 2002.

[75] M. S. Jog, Y. Kubota, C. I. Connolly, V. Hillegaart, and A. M. Graybiel. Building neural representations of habits. *Science*, 286(5445):1745, 1999.

[76] R. S. Johansson, G. Westling, A. Bäckström, and J. R. Flanagan. Eye-hand coordination in object manipulation. *Journal of Neuroscience*, 21(17):6917–32, 2001.

[77] K. E. Jones, A. F. C. Hamilton, and D. M. Wolpert. Sources of signal-dependent noise during isometric force production. *Journal of Neurophysiology*, 88(3):1533–1544, 2002.

[78] M. Jordan. *Computational aspects of motor control and motor learning.* Handbook of Perception and Action: Motor Skills. Academic Press, New York, 1996.

[79] M. Jueptner, C. D. Frith, D. J. Brooks, R. S. Frackowiak, and R. E. Passingham. Anatomy of motor learning. ii. subcortical structures and learning by trial and error. *Journal of Neurophysiology*, 77(3):1325–37, 1997.

[80] M. Jueptner and C. Weiller. A review of differences between basal ganglia and cerebellar control of movements as revealed by functional imaging studies. *Brain*, 121:1437–1449, 1998.

[81] S. Kakade. A natural policy gradient. *Advances in Neural Information Processing Systems*, 14, 2002.

[82] J. F. Kalaska. Parietal cortex area 5: a neuronal representation of movement kinematics for kinæsthetic perception and movement control. In J. Paillard, editor, *Brain and Space*. Oxford Univ. Press Oxford, UK, 1991.

[83] J. F. Kalaska, R. Caminiti, and A. P. Georgopoulos. Cortical mechanisms related to the direction of two-dimensional arm movements: relations in parietal area 5 and comparison with motor cortex. *Experimental Brain Research*, 51(2):247–260, 1983.

[84] J. F. Kalaska and D. J. Crammond. Cerebral cortical mechanisms of reaching movements. *Science*, 255(5051):1517–1523, 1992.

[85] R.E. Kalman. A new approach to linear filtering and prediction problems. *ASME - Journal of Basic Engineering*, 82:35–45, 1960.

[86] E. R. Kandel, J. H. Schwartz, and T. M. Jessell. *Principles of Neural Science*. McGraw-Hill/Appleton & Lange, 2000.

[87] A. Karni, G. Meyer, C. Rey-Hipolito, P. Jezzard, M. M. Adams, R. Turner, and L. G. Ungerleider. The acquisition of skilled motor performance: Fast and slow experience-driven changes in primary motor cortex. *Proceedings of the National Academy of Sciences*, 95(3):861, 1998.

[88] M. Kawato. Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, 9(6):718–727, 1999.

[89] M. Kawato and K. Samejima. Efficient reinforcement learning: computational theories, neuroscience and robotics. *Current Opinion in Neurobiology*, 17(2):205–212, 2007.

[90] S. W. Keele and M. I. Posner. Processing of visual feedback in rapid movements. *Journal of Experimental Psychology*, 77(1):155–8, 1968.

[91] S. W. Kennerley, K. Sakai, and M. F. Rushworth. Organization of action sequences and the role of the pre-sma. *Journal of Neurophysiology*, 91(2):978–93, 2004.

[92] S. Kitazawa, T. Kohno, and T. Uka. Effects of delayed visual information on the rate and amount of prism adaptation in the human. *Journal of Neuroscience*, 15(11):7644–52, 1995.

[93] S. Kitazawa and P.B. Yin. Prism adaptation with delayed visual error signals in the monkey. *Experimental Brain Research*, 144(2), 2002.

[94] A. H. Klopf. *The hedonistic neuron: a theory of memory, learning, and intelligence*. Hemisphere Washington, DC, 1982.

[95] Y. Kobayashi, K. Kawano, A. Takemura, Y. Inoue, T. Kitama, H. Gomi, and M. Kawato. Temporal firing patterns of purkinje cells in the cerebellar ventral paraflocculus during ocular following responses in monkeys ii. complex spikes. *Journal of Neurophysiology*, 80(2):832–848, 1998.

[96] K. P. Körding and D. M. Wolpert. Bayesian integration in sensorimotor learning. *Nature*, 427(6971):244–7, 2004.

[97] K. P. Körding and D. M. Wolpert. Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, 10(7):319–26, 2006.

[98] M. Kositsky and A. G. Barto. The emergence of multiple movement units in the presence of noise and feedback delay. *Advances in Neural Information Processing Systems 14: Proceedings of the 2002 Conference*, 2002.

[99] M. F. Land. Predictable eye-head coordination during driving. *Nature*, 359(6393):318–20, 1992.

[100] M. F. Land and D. N. Lee. Where we look when we steer. *Nature*, 369(6483):742–4, 1994.

[101] M. F. Land and P. McLeod. From eye movements to actions: how batsmen hit the ball. *Nature Neuroscience*, 3(12):1340–5, 2000.

[102] D. M. Levi, S. A. Klein, and A. P. Aitsebaomo. Vernier acuity, crowding and cortical magnification. *Vision Research*, 25(7):963–77, 1985.

[103] D. Liu and E. Todorov. Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. *Journal of Neuroscience*, 27(35):9354–68, 2007.

[104] X. Lu, O. Hikosaka, and S. Miyachi. Role of monkey cerebellar nuclei in skill for sequential movement. *Journal of Neurophysiology*, 79(5):2245–54, 1998.

[105] T. Matsubara, J. Morimoto, J. Nakanishi, M. Sato, and K. Doya. Learning sensory feedback to cpg with policy gradient for biped locomotion. *Robotics and Automation, 2005. Proceedings of the 2005 IEEE International Conference on*, pages 4164–4169, 2005.

[106] Y. Matsuzaka, H. Aizawa, and J. Tanji. A motor area rostral to the supplementary motor area (presupplementary motor area) in the monkey: neuronal activity during a learned motor task. *Journal of Neurophysiology*, 68(3):653–662, 1992.

[107] P. S. Maybeck, J. D. A. John, C. H. Houpis, H. B. Kepler, V. Rehg, J. Klosterman, C. T. Triscari Jr, M. R. W. Hitzelberger, D. G. Shankland, and M. T. C. Harrington. Stochastic models, estimating, and control. *Management Quarterly*, 8:12–14, 1982.

[108] S. M. McClure, K. M. Ericson, D. I. Laibson, G. Loewenstein, and J. D. Cohen. Time discounting for primary rewards. *Journal of Neuroscience*, 27(21):5796–804, 2007.

[109] B. Mehta and S. Schaal. Forward models in visuomotor control. *Journal of Neurophysiology*, 88(2):942–53, 2002.

[110] J. Mena-Segovia, J. P. Bolam, and P. J. Magill. Pedunculopontine nucleus and basal ganglia: distant relatives or part of the same family? *Trends in Neurosciences*, 27(10):585–588, 2004.

[111] R. C. Miall and J. K. Jackson. Adaptation to visual feedback delays in manual tracking: evidence against the smith predictor model of human visually guided action. *Experimental Brain Research*, 172(1):77–84, 2006.

[112] R. C. Miall, D. J. Weir, and J. F. Stein. Manual tracking of visual targets by trained monkeys. *Behavioural Brain Research*, 20(2):185–201, 1986.

[113] R. C. Miall, D. J. Weir, D. M. Wolpert, and J. F. Stein. Is the cerebellum a smith predictor. *Journal of Motor Behavior*, 25(3):203–216, 1993.

[114] F. A. Middleton and P. L. Strick. Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain Research Reviews*, 31(2-3):236–50, 2000.

[115] M. S. Milak, Y. Shimansky, V. Bracha, and J. R. Bloedel. Effects of inactivating individual cerebellar nuclei on the performance and retention of an operantly conditioned forelimb movement. *Journal of Neurophysiology*, 78(2):939–959, 1997.

[116] S. Miyachi, O. Hikosaka, and X. Lu. Differential activation of monkey striatal neurons in the early and late stages of procedural learning. *Experimental Brain Research*, 146(1):122–6, 2002.

[117] S. Miyachi, O. Hikosaka, K. Miyashita, Z. Kárádi, and M. K. Rand. Differential roles of monkey striatum in learning of sequential hand movement. *Experimental Brain Research*, 115(1):1–5, 1997.

[118] K. Miyashita, M. K. Rand, S. Miyachi, and O. Hikosaka. Anticipatory saccades in sequential procedural learning in monkeys. *Journal of Neurophysiology*, 76(2):1361–6, 1996.

[119] P. R. Montague, P. Dayan, and T. J. Sejnowski. A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of Neuroscience*, 16(5):1936–47, 1996.

[120] J. Morimoto and K. Doya. Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning. *Robotics and Autonomous Systems*, 36(1):37–51, 2001.

[121] C. E. Myers, D. Shohamy, M. A. Gluck, S. Grossman, A. Kluger, S. Ferris, J. Golomb, G. Schnirman, and R. Schwartz. Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *Journal of Cognitive Neuroscience*, 15(2):185–193, 2003.

[122] J. Najemnik and W. S. Geisler. Optimal eye movement strategies in visual search. *Nature*, 434(7031):387–91, 2005.

[123] H. Nakahara, K. Doya, and O. Hikosaka. Parallel cortico-basal ganglia mechanisms for acquisition and execution of visuomotor sequences - a computational approach. *Journal of Cognitive Neuroscience*, 13(5):626–47, 2001.

[124] K. Nakamura and C. L. Colby. Visual, saccade-related, and cognitive activation of single neurons in monkey extrastriate area v3a. *Journal of Neurophysiology*, 84(2):677–692, 2000.

[125] K. Nakamura and C. L. Colby. Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proceedings of the National Academy of Sciences*, 99(6):4026–31, 2002.

[126] K. Nakamura, K. Sakai, and O. Hikosaka. Neuronal activity in medial frontal cortex during learning of sequential procedures. *Journal of Neurophysiology*, 80(5):2671–87, 1998.

[127] K. Nakamura, K. Sakai, and O. Hikosaka. Effects of local inactivation of monkey medial frontal cortex in learning of sequential procedures. *Journal of Neurophysiology*, 82(2):1063–8, 1999.

[128] P. D. Nixon and R. E. Passingham. The cerebellum and cognition: cerebellar lesions impair sequence learning but not conditional visuomotor learning in monkeys. *Neuropsychologia*, 38(7):1054–1072, 2000.

[129] K. Ogawa, T. Inui, and T. Sugio. Neural correlates of state estimation in visually guided movements: an event-related fmri study. *Cortex*, 43(3):289–300, 2007.

[130] R. Osu, N. Kamimura, H. Iwasaki, E. Nakano, C. M. Harris, Y. Wada, and M. Kawato. Optimal impedance control for task achievement in the presence of signal-dependent noise. *Journal of Neurophysiology*, 92(2):1199–1215, 2004.

[131] F. P. Ottes, J. A. Van Gisbergen, and J. J. Eggermont. Metrics of saccade responses to visual double stimuli: two different modes. *Vision Research*, 24(10):1169–79, 1984.

[132] I. P. Pavlov. *Conditioned Reflexes*. Dover Publications, 2003.

[133] J. Peters and S. Schaal. Applying the episodic natural actor-critic architecture to motor primitive learning. In *Proceedings of the 2007 European symposium on Artificial Neural Networks (ESANN)*, Bruges, Belgium, 2007.

[134] J. Peters, S. Vijayakumar, and S. Schaal. Natural actor-critic. *Proceedings of the Sixteenth European Conference on Machine Learning (ECML05)*, 3720, 2005.

[135] E.S. Petersen, H. van Mier, A.J. Fiez, and E.M. Raichle. The effect of practice on the functional anatomy of task performance. *Proceedings of the National Academy of Sciences*, 95:853–860, 1998.

[136] N. Picard and P. L. Strick. Motor areas of the medial wall: A review of their location and functional activation. *Cerebral Cortex*, 6(3):342–353, 2003.

[137] A. Pouget, P. Dayan, and R. S. Zemel. Inference and computation with population codes. *Annual Review of Neuroscience*, 26:381–410, 2003.

[138] C. Prablanc, J. F. Echallier, E. Komilis, and M. Jeannerod. Optimal response of eye and hand motor systems in pointing at a visual target. i. spatio-temporal characteristics of eye and hand movements and their relationships when varying the amount of visual information. *Biological Cybernetics*, 35(2):113–24, 1979.

[139] L. W. Renninger, P. Verghese, and J. Coughlan. Where to look next? eye movements reduce local uncertainty. *Journal of Vision*, 7(3):6, 2007.

[140] Y. Rossetti, B. Tadary, and C. Prablanc. Optimal contributions of head and eye positions to spatial accuracy in man tested by visually directed pointing. *Experimental Brain Research*, 97(3):487–496, 1994.

[141] G. S. Russo, D. A. Backus, S. Ye, and M. D. Crutcher. Neural activity in monkey dorsal and ventral cingulate motor areas: Comparison with the supplementary motor area. *Journal of Neurophysiology*, 88(5):2612–2629, 2002.

[142] U. Sailer, J. R. Flanagan, and R. S. Johansson. Eye-hand coordination during learning of a novel visuomotor task. *Journal of Neuroscience*, 25(39):8833–42, 2005.

[143] K. Sakai, K. Kitaguchi, and O. Hikosaka. Chunking during human visuomotor sequence learning. *Experimental Brain Research*, 152(2):229–42, 2003.

[144] B. Sakitt and H.B. Barlow. A model for the economical encoding of the visual image in cerebral cortex. *Biological Cybernetics*, 43(2):97–108, 1982.

[145] K. Samejima, Y. Ueda, K. Doya, and M. Kimura. Representation of action-specific reward values in the striatum. *Science*, 310(5752):1337–1340, 2005.

[146] M. T. Schmolesky, Y. Wang, D. P. Hanes, K. G. Thompson, S. Leutgeb, J. D. Schall, and A. G. Leventhal. Signal timing across the macaque visual system. *Journal of Neurophysiology*, 79(6):3272–3278, 1998.

[147] W. Schultz, P. Apicella, and T. Ljungberg. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *Journal of Neuroscience*, 13(3):900–13, 1993.

[148] W. Schultz, P. Dayan, and P. R. Montague. A neural substrate of prediction and reward. *Science*, 275(5306):1593–9, 1997.

[149] N. Schweighofer, K. Shishida, C. E. Han, Y. Okamoto, S. C. Tanaka, S. Yamawaki, and K. Doya. Humans can adopt optimal discounting strategy under real-time constraints. *PLoS Computational Biology*, 2(11):e152, 2006.

[150] M. I. Sereno, A. M. Dale, J. B. Reppas, K. K. Kwong, J. W. Belliveau, T. J. Brady, B. R. Rosen, and R. B. Tootell. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268(5212):889–893, 1995.

[151] O.J.M. Smith. A controller to overcome dead time. *ISA Journal*, 6:28–33, 1959.

[152] R. E. Suri and W. Schultz. Temporal difference model reproduces anticipatory neural activity. *Neural Computation*, 13(4):841–62, 2001.

[153] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.

[154] R. S. Sutton, D. McAllester, S. Singh, and Y. Mansour. Policy gradient methods for reinforcement learning with function approximation. *Advances in Neural Information Processing Systems*, 12:10571063, 2000.

[155] S. C. Tanaka, K. Doya, G. Okada, K. Ueda, Y. Okamoto, and S. Yamawaki. Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neuroscience*, 7(8):887–93, 2004.

[156] J. Tanji. The supplementary motor area in the cerebral cortex. *Neuroscience Research*, 19(3):251–68, 1994.

[157] J. Tanji. Sequential organization of multiple movements: Involvement of cortical motor areas. *Annual Review of Neuroscience*, 24(1):631–651, 2001.

[158] Y. Terao, N. E. Andersson, J. R. Flanagan, and R. S. Johansson. Engagement of gaze in capturing targets for future sequential manual actions. *Journal of Neurophysiology*, 88(4):1716–25, 2002.

[159] G. Tesauro. Td-gammon: a self-teaching backgammon program. *Applications of Neural Networks*, 1995.

[160] E. L. Thorndike. *Animal Intelligence*. Thoemmes Press [ua], 1998.

[161] E. Todorov. Optimality principles in sensorimotor control. *Nature Neuroscience*, 7(9):907–15, 2004.

[162] E. Todorov. Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. *Neural Computation*, 17(5):1084–108, 2005.

[163] E. Todorov and M. I. Jordan. Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5(11):1226–35, 2002.

[164] I. Toni, M. Krams, R. Turner, and R. E. Passingham. The time course of changes during motor sequence learning: A whole-brain fmri study. *Neuroimage*, 8(1):50–61, 1998.

[165] I. Toni and R. E. Passingham. Prefrontal-basal ganglia pathways are involved in the learning of arbitrary visuomotor associations: a pet study. *Experimental Brain Research*, 127(1):19–32, 1999.

[166] M. M. Umeno and M. E. Goldberg. Spatial processing in the monkey frontal eye field. i. predictive visual responses. *Journal of Neurophysiology*, 78(3):1373–83, 1997.

[167] M. M. Umeno and M. E. Goldberg. Spatial processing in the monkey frontal eye field. ii. memory responses. *Journal of Neurophysiology*, 86(5):2344–52, 2001.

[168] Y. Uno, M. Kawato, and R. Suzuki. Formation and control of optimal trajectory in human multijoint arm movement. *Biological Cybernetics*, 61(2):89–101, 1989.

[169] J. N. Vickers. Gaze control in putting. *Perception*, 21(1):117–32, 1992.

[170] J. N. Vickers. Visual control when aiming at a far target. *Journal of Experimental Psychology: Human Perception and Performance*, 22(2):342–54, 1996.

[171] M. F. Walker, E. J. Fitzgibbon, and M. E. Goldberg. Neurons in the monkey superior colliculus predict the visual result of impending saccadic eye movements. *Journal of Neurophysiology*, 73(5):1988–2003, 1995.

[172] Y. Weiss and D.J. Fleet. Velocity likelihoods in biological and machine vision. In R. Rao, B. Olshausen, and M.S. Lewicki, editors, *Statistical theories of the cortex*, pages 77–96. MIT Press, Cambridge, MA, 2002.

[173] F.W. Weymouth. Visual sensory units and the minimal angle of resolution. *American Journal of Ophtalmology*, 46-2:102–13, 1958.

[174] R. J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Reinforcement Learning*, 8:229–256, 1992.

[175] D. M. Wolpert, R.C. Miall, and M. Kawato. Internal models in the cerebellum. *Trends in Cognitive Sciences*, 2(9):338–346, 1998.