



PERGAMON

Neural Networks 11 (1998) 1317–1329

Neural  
Networks

1998 Special Issue

# Multiple paired forward and inverse models for motor control

D.M. Wolpert<sup>a,\*</sup>, M. Kawato<sup>b</sup>

<sup>a</sup>*Sobell Department of Neurophysiology, Institute of Neurology, Queen Square, London WC1N 3BG, UK*

<sup>b</sup>*ATR Human Information Processing Research Laboratories and Dynamic Brain Project, ERATO, JST, Kyoto, Japan*

Received and accepted 30 April 1998

## Abstract

Humans demonstrate a remarkable ability to generate accurate and appropriate motor behavior under many different and often uncertain environmental conditions. In this paper, we propose a modular approach to such motor learning and control. We review the behavioral evidence and benefits of modularity, and propose a new architecture based on multiple pairs of inverse (controller) and forward (predictor) models. Within each pair, the inverse and forward models are tightly coupled both during their acquisition, through motor learning, and use, during which the forward models determine the contribution of each inverse model's output to the final motor command. This architecture can simultaneously learn the multiple inverse models necessary for control as well as how to select the inverse models appropriate for a given environment. Finally, we describe specific predictions of the model, which can be tested experimentally. © 1998 Elsevier Science Ltd. All rights reserved.

*Keywords:* Motor control; Modularity; Internal models; Motor learning; Contextual prediction

## 1. Introduction

Humans exhibit an enormous repertoire of motor behavior which enables us to interact with many different objects under a variety of different environments. The ability to perform in such a varying and often uncertain environment is a feature which is conspicuously absent from most robotic control, as robots tend to be designed to operate within rather limited environmental situations. In general, the problem of control can be considered as the computational process of determining the input to some system we wish to control which will achieve some desired output. In human motor control, the problem might be to select the input, i.e. motor command, to achieve some required output, i.e. desired sensory feedback. If we consider an example of lifting a can to one's lips, it may be that the desired output at a specific time is a particular acceleration of the hand as judged by sensory feedback. However, the motor command needed to achieve this acceleration will depend on many variables, both internal and external to the body. Clearly, the motor command depends on the state of the arm, i.e. its joint angles and angular velocities. The dynamic equations governing the system also depend on some relatively unvarying parameters, e.g. masses, moments of inertia,

and center of masses of the upper arm and forearm. However, these parameters specific to the arm are insufficient to determine the motor command necessary to produce the desired hand acceleration; knowledge of the interactions with the outside world must also be known. For example, the geometry and inertial properties of the can will alter the arm's dynamics. More global environmental conditions also contribute to the dynamics, e.g. the orientation of the body relative to gravity and the angular acceleration of the torso about the body. As these parameters are not directly linked to the quantities we can measure about the arm, we will consider them as representing the context of the movement. As the context of the movement alters the input–output relationship of the system under control, the motor command must be tailored to take account of the current context.

Considering the number of objects and environments, and their possible combinations, which can influence the dynamics of the arm (let alone the rest of the body), the motor control system must be capable of providing appropriate motor commands for the multitude of distinct contexts that are likely to be experienced. Given the abundance of contexts within which we must act, there are two qualitatively distinct strategies to motor control and learning. The first is to use a single controller which uses all the contextual information in an attempt to produce an appropriate control signal. However, such a controller would demand enormous

\* Corresponding author. Fax: 0171-813 3107; e-mail: wolpert@hera.ucl.ac.uk

complexity to allow for all possible scenarios. If this controller were unable to encapsulate all the possible contexts, it would need to adapt every time the context of the movement changed before it could produce appropriate motor commands—this would produce transient and possibly large performance errors. Alternatively, a modular approach can be used in which multiple controllers co-exist, with each controller suitable for one or a small set of contexts. Based on an estimate of the current context, some of the controllers could be activated to generate the appropriate motor command. Under such a modular strategy, there would need to be a context identification process which could choose the appropriate controllers from the set of all possible controllers.

In this paper, we focus on how we learn to produce effective motor control under a variety of contexts. The approach we take is to employ a modular system to tackle the problems of motor learning and control. We will first describe the type of models which must be learned and the benefits which accrue from such modularity. Experimental evidence that supports the use of multiple functionally discrete controllers in humans is briefly reviewed. We then focus on how such modular controllers can be learned and selected, and propose a neural architecture based on multiple paired forward–inverse models. Finally, we describe specific predictions of the model which can be tested experimentally.

## 2. Modularity in motor control

In this section, we describe the type of models which must be learned in motor control, i.e. internal forward and inverse models. We then describe the potential benefits of learning these models in a modular fashion.

### 2.1. Internal models

The notion of an internal model, a system which mimics the behavior of a natural process, has emerged as an important theoretical concept in motor control. There are two varieties of internal model, forward and inverse models. Forward models capture the forward or causal relationship between inputs to the system, e.g. the arm, and the outputs (Ito, 1970; Kawato et al., 1987; Jordan, 1995). A forward dynamic model of the arm, for example predicts the next state (e.g. position and velocity) given the current state and motor command. Such models have been proposed to be used in motor learning (Sutton and Barto, 1981; Jordan and Rumelhart, 1992), state estimation (Wolpert et al., 1995b) and motor control (Ito, 1984; Miall et al., 1993; Wolpert, 1997). In contrast, inverse models invert the system by providing the motor command which will cause a desired change in state. Inverse models are, therefore, well suited to act as controllers as they can provide the motor command necessary to achieve some desired state

transition. Even control strategies, such as feedback control, which do not explicitly invoke an inverse model, can be thought of as implicitly constructing an inverse model.

As both forward and inverse models depend on the dynamics of the motor system, which change throughout life and under different contextual conditions, these models must be adaptable. While forward models can be learned relatively straightforwardly by supervised learning, comparing the predicted consequences of an action to the actual result, inverse models prove more problematic. If the correct motor command was known, which could provide an appropriate supervised error signal, then there would be no need for the inverse model. Three main approaches have been used to adapt such inverse models—direct inverse modeling (Miller, 1987; Kuperstein, 1988), distal supervised learning (Jordan and Rumelhart, 1992) and feedback-error-learning (Kawato, 1990). The latter two models both rely on the ability to convert errors in the actual trajectory into errors in the motor command. They, unlike the direct approach, are able to acquire an accurate inverse model even for redundant systems.

### 2.2. Benefits of modularity

While forward and inverse models could be learned by a single module, there are three potential benefits in employing a modular approach. First, the world is essentially modular, in that we interact with multiple qualitatively different objects and environments. By using multiple inverse models, each of which might capture the motor commands necessary when acting with a particular object or within a particular environment, we can achieve an efficient coding of the world. In other words, the large set of environmental conditions in which we are required to generate movement requires multiple behaviors or sets of motor commands, each embodied within a module.

Second, the use of a modular system allows individual modules to participate in motor learning without affecting the motor behaviors already learned by other modules. Such modularity can therefore reduce temporal crosstalk, thereby both speeding up motor learning while retaining previously learned behaviors.

Third, many situations which we encounter are derived from combinations of previously experienced contexts, e.g. novel conjunctions of manipulated objects and environments. By modulating the contribution of the outputs of the inverse models to the final motor command, an enormous repertoire of behaviors can be generated. With as few as 32 inverse models, in which the outputs of each model either contribute or do not contribute to the final motor command, we have  $2^{32}$  or  $10^{10}$  behaviors—sufficient for a new behavior for every second of one's life. Therefore, multiple internal models can be regarded conceptually as motor primitives, which are the building blocks used to construct intricate motor behaviors with an enormous vocabulary.

### 3. Experimental evidence for modularity

#### 3.1. Do multiple independent controllers exist within the CNS?

Studies of motor adaptation have suggested that we are able to learn multiple controllers and switch between them based on context. In general, when subjects are brought into the laboratory to undergo a motor learning task in which they must adapt to a visual or dynamic perturbation, they can take many movements to adapt (Welch, 1986; Shadmehr and Mussa-Ivaldi, 1994). Although the time course of adaptation can extend over hours, on removal of the perturbation, de-adaptation is often very rapid. In some cases, just removing the subject from the experimental equipment is enough to restore pre-perturbation behavior. Such asymmetry between learning and “unlearning” suggests that learning may represent adaptation of a new module, whereas de-adaptation represents the switching back to a previously learned stable module.

In agreement with this interpretation is work on re-adaptation. On repeated presentation of a visual (McGonigle and Flook, 1978; Welch et al., 1993) or dynamic perturbation (Brashers-Krug et al., 1996), subjects adapt increasingly rapidly. This suggests that the retained module for the adapted state is not destroyed by de-adaptation and, moreover, it can be quickly switched on again in response to its introduction.

While the adaptation or switching already described can be attributed to performance errors or knowledge of the consequences of one’s actions, there is evidence that a switching process is also in operation dependent on purely sensory components of the context. Several studies have examined the degree to which two different perturbations can be learned and switched between. In these experiments, one or more perturbations are introduced, and the nature of the perturbation is contingent either on the configuration of the body or on some other sensory cue. For example, when subjects are repeatedly exposed to a prismatic displacement induced by wearing prism goggles, they eventually show adaptive changes cued by the feel of the prism glasses without any prism lenses (Kravitz, 1972; Welch, 1971; Martin et al., 1996). Similarly, context-dependent adaptation can also be seen if cued by gaze direction (Kohler, 1950; Hay and Pick, 1966; Shelhamer et al., 1991), body orientation (Baker et al., 1987), arm configuration (Gandolfo et al., 1996) or an auditory tone (Kravitz and Yaffe, 1972). These studies suggest that subjects can switch immediately between two learned behaviors based on the context.

#### 3.2. Can we combine the output of independent controllers?

While the studies described in the previous section suggest that multiple modules can be learned, they do not address whether two modules can be appropriately activated at the same time. Data for mixing of two new learned

modules based on prism work (Ghahramani and Wolpert, 1997) suggest a specific way that multiple modules are integrated. Using a virtual reality system, a single visual target location was remapped to two different hand positions depending on the starting location of the movement. Such a perturbation creates a conflict in the visuomotor map which captures the (normally one-to-one) relation between visually perceived and actual hand locations. One way to resolve this conflict is to develop two separate visuomotor maps, each appropriate for one of the two starting locations. A separate mechanism could then combine, based on the starting location of the movement, the outputs of the two visuomotor maps. The internal structure of the system was probed by investigating the generalization properties in response to novel inputs, which in this case are the starting locations on which it has not been trained. As predicted by a modular architecture, subjects were able to learn both conflicting mappings, and to interpolate smoothly from one visuomotor map to the other as the starting location was varied. This provides evidence that two modules’ outputs can be mixed, as the context is varied between the contexts under which each was learned.

### 4. General methodology for multiple modules

Based on the benefits of a modular approach and the experimental evidence for modularity, we propose that the problem of motor learning and control is best solved using multiple controllers, i.e. inverse models. At any one time, one or a subset of these inverse models will contribute to the final motor command. Such a system raises two fundamental computational problems. First, given a set of inverse models which appropriately cover the set of contexts which might be experienced, how is the correct subset of inverse models selected for the particular current context—the module selection problem. Second, how are the set of inverse models learned to cover all the contexts which might be experienced—the module learning problem.

#### 4.1. Module selection problem

We first consider the module selection problem, assuming that we already have a set of learned inverse models. How can the outputs of the inverse models be switched on and off appropriately in response to different behavioral contexts so as to generate a coordinated final motor command? From human psychophysical data, we know that such a switching process must be driven by two distinct processes. The first process is a feedforward adjustment in motor commands based on purely sensory signals, e.g. the perceived size of an object or whether a cup looks full or empty. Such adjustments are made without regard to the consequences of the action and therefore may need to be altered by a second switching process, based on feedback of the outcome of a movement. For example, on picking up a can which appears

full, feedforward switching may activate modules responsible for generating a large motor impulse necessary to raise the can to our lips. However, feedback processes, based on contact with the can, can indicate that the can is in fact empty, thereby switching off the inverse model for a heavy can and activating an inverse model appropriate for the arm in contact with a light can.

The feedforward selection process must be driven by sensory cues alone, and learned from experience to switch between the controller modules. The feedback switching could be driven by a variety of signals. A natural signal to consider would be the discrepancy between desired and actual state, i.e. the performance error. The performance error reflects the performance of the currently active module and therefore its suitability for the current environment. This signal could be used to determine, based on a performance threshold, when it might be appropriate to switch controls. However, it cannot provide information as to which of the other controllers is appropriate to activate. This is because at any one time only the active controllers performance can be assessed and the inactive controllers performance error could be assessed only after they are activated.

#### 4.2. Module learning problem

As the parameters of the motor system change drastically during growth and under different contexts, inverse models are not expected to be genetically inherited. The problem therefore arises as to how they are acquired from experience. Such learning must be able to divide up the control into appropriate modules which can be recombined to generate behaviors. The learning must also be robust to learning new contexts, in that previously learned models must be relatively stable while other modules are learned.

### 5. Multiple paired forward–inverse model

In this section, we propose a model which can solve the module learning and selection problems in a computationally coherent manner from a single principle. The basic idea of the model is that multiple inverse models exist to control the system and each is augmented with a corresponding forward model. The brain therefore contains multiple pairs of corresponding forward and inverse models. Within each pair, the inverse and forward internal models are tightly coupled both during their acquisition through motor learning, and use through gating of the inverse models' outputs dependent on the behavioral context. The key to this model is the responsibility signals which reflect, at any given time, the degree to which each pair of forward and inverse models should be responsible for controlling the current behavior. This responsibility signal is derived from the combination of two processes. The first process uses the forward model's predictions. As the forward models capture distinct

dynamical behaviours of the motor system, their prediction errors can be used during movement to determine in which context the motor system is acting. The second process, the responsibility predictors, use sensory contextual cues to predict the responsibility of the module and can therefore select controllers prior to movement initiation in a feedforward manner. This responsibility signal both couples the inverse and forward model pairs, guides learning in each pair of the inverse and forward models, and gates the contribution of each inverse model's output to the final output.

Here, we consider a motor system which we wish to control acted upon by motor command  $u_t$  at time  $t$  (for simplicity we will consider a discrete time system, although an equivalent continuous time formulation could be considered). The resulting (actual) movement trajectory,  $x_t$ , is determined by Eq. (1), which describes the causal relationship between the motor command and the movement governed by the function  $f$  which describes the forward dynamics of the motor apparatus.

$$x_{t+1} = f(x_t, u_t) \quad (1)$$

The aim of the control is to produce a system which can generate an appropriate motor command  $u_t$  given the desired state,  $x_{t+1}^*$ , (in general  $g$  need not be unique or exist for all  $x_{t+1}^*$ ; such conditions do not affect the model's formulation):

$$u_t = g(x_{t+1}^*, x_t) \quad (2)$$

so that  $g$  and  $f$  have an inverse relationship

$$x_{t+1}^* = f(x_t, g(x_{t+1}^*, x_t)). \quad (3)$$

However, we assume that the dynamics of the system  $f$  are not fixed over time but can take on a possibly infinite number of different forms. These different forms correspond to the context of the movement and include such factors as interactions with objects or changes in the environment. This can either be parameterized by assuming there is a set of system dynamics  $f_i$  with  $i = 1, 2, \dots, n$  or by including a context parameter  $c$  as part of the dynamics

$$x_{t+1} = f(x_t, u_t, c_t) \quad (4)$$

where  $c_t$  encapsulates the context of the movement at time  $t$ . The aim of the overall controller is to learn to control the system under different and unknown contexts.

In the next two sections, we describe how switching can be guided during movement, and in Section 5.3, we describe how this mechanism can be modified to include pre-movement switching.

#### 5.1. Multiple forward models—dividing up experience

Central to the multiple paired forward–inverse model is the notion of dividing up experience using predictive forward models. Under different contexts,  $c_t$ , the consequence of performing the same motor command  $u_t$  from the same

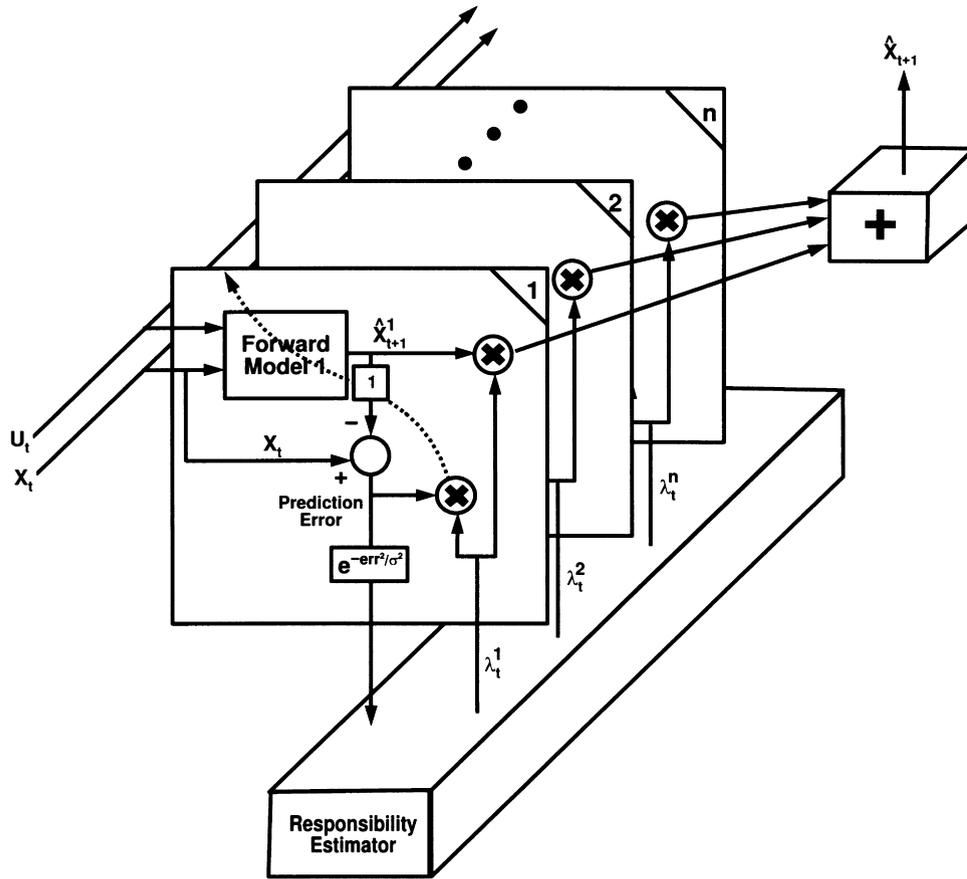


Fig. 1. Multiple forward models dividing up experience. The forward model component of the full model is shown. Each forward model predicts the next state based on the motor command and current state. This prediction is delayed and compared with the actual next state. The prediction errors are used to assign responsibility to each module which determines the learning (dotted line) as well as the contribution each model's prediction makes to the final prediction. The responsibility estimator normalizes the transformed error signals  $e^{-err^2/\sigma^2}$ . The responsibilities are therefore produced using the soft-max function on the prediction errors.

state  $x_t$  will be different. Therefore, a single predictor which has access only to the state and motor command will be unable to predict the consequences of performed actions under different contexts. We propose that multiple predictors exist with at least one able to provide an accurate prediction of the next state of the system given any context.

To generate such a set of predictors, we propose that the forward models learn to partition experience by competitive self-supervised learning (Fig. 1). We consider a set of undifferentiated forward models which each receive the current state and motor command as inputs. The outputs of the forward models,  $\hat{x}_{t+1}^1, \hat{x}_{t+1}^2, \dots, \hat{x}_{t+1}^n$ , represent the predictions, at time  $t$ , of the next state made by the 1st, 2nd, ..., and  $n$ -th forward model, respectively. Therefore, each forward model attempts to predict the next state of the system given the current state and motor command. The output of the  $i$ -th forward model at time  $t$  is given as  $\hat{x}_{t+1}^i$  where:

$$\hat{x}_{t+1}^i = \phi(w_t^i, x_t, u_t) \tag{5}$$

where  $w_t^i$  are the parameters of a function approximator  $\phi$  (e.g. neural network weights) used to model the forward dynamics. These predicted next states could then be

compared to the actual next state to provide a signal suitable for supervised learning. To ensure that the forward models learn to competitively divide up the experienced relationships, the training signal for each forward model is gated by a responsibility signal  $\lambda_t^i$ . This responsibility signal represents the extent to which each forward model presently accounts for the behavior of the system. Based on the prediction errors of the forward models, the responsibility signal,  $\lambda_t^i$ , for the  $i$ -th forward-inverse model pair is calculated by the soft-max

$$\lambda_t^i = \frac{e^{-|x_t - \hat{x}_t^i|^2/\sigma^2}}{\sum_{j=1}^n e^{-|x_t - \hat{x}_t^j|^2/\sigma^2}} \tag{6}$$

where  $x_t$  is the true state of the system and  $\sigma$  is a scaling constant. The soft-max transforms the errors using the exponential function and then normalizes these values, within the responsibility estimator (Fig. 1), across the modules so that the responsibilities have the property that they lie between 0 and 1, and the responsibility summed over the models is 1. These responsibilities can be thought of as a measure of the

probability that the  $i$ -th module captures the behavior. Those forward models which capture the current behavior, and therefore have low errors, will have high responsibilities. The responsibilities are then used to control the learning within the forward models, with those models with high responsibilities receiving proportionally more of their error signals than modules with low responsibility. The final prediction is made by modulating the contribution of each forward models' output to the final prediction of state  $\hat{x}_{t+1}$

$$\hat{x}_{t+1} = \sum_{i=1}^n \lambda_i \hat{x}_{t+1}^i = \sum_{i=1}^n \lambda_i \phi(w_i^i, x_t, u_t) \quad (7)$$

Based on this prediction, a gradient descent learning rule is given by

$$\Delta w_t^i = \epsilon \lambda_t^i \frac{d\phi_i}{dw_t^i}(x_t - \hat{x}_t^i) = \epsilon \frac{d\hat{x}_t^i}{dw_t^i} \lambda_t^i (x_t - \hat{x}_t^i) \quad (8)$$

Over time, the forward model will learn to divide up the system dynamics experienced, and the responsibilities will reflect the extent to which each forward model captures the current dynamics of the system.

## 5.2. Multiple inverse models—controlling the system

Based on the multiple forward models' ability to divide up experience, we suggest that for each behavior captured by a forward model we would wish to learn a control strategy, i.e. an inverse model. Therefore, for each forward model there would be a paired inverse model. Each inverse model would have as input the desired next state  $x_{t+1}^*$  and would produce a motor command  $u_t^i$  as output (Fig. 2). Therefore, each inverse model produces a motor output with  $u_t^1, u_t^2, \dots, u_t^n$  denoting the outputs from the 1st, 2nd, ...,  $n$ -th inverse models. The aim is that each inverse model learns to provide suitable control signals under the context for which its paired forward model provides good predictions. The  $i$ -th inverse model would produce a motor output  $u_t^i$  based on the desired state  $x_t^*$ :

$$u_t^i = \psi(\alpha_i^i, x_{t+1}^*) \quad (9)$$

where  $\alpha_i^i$  are the parameters of some function approximator  $\psi$ . For simplicity, here we assume that the motor command of each module is calculated purely in a feedforward fashion from the desired trajectory, but it is straightforward to extend this to also include the feedback of the current

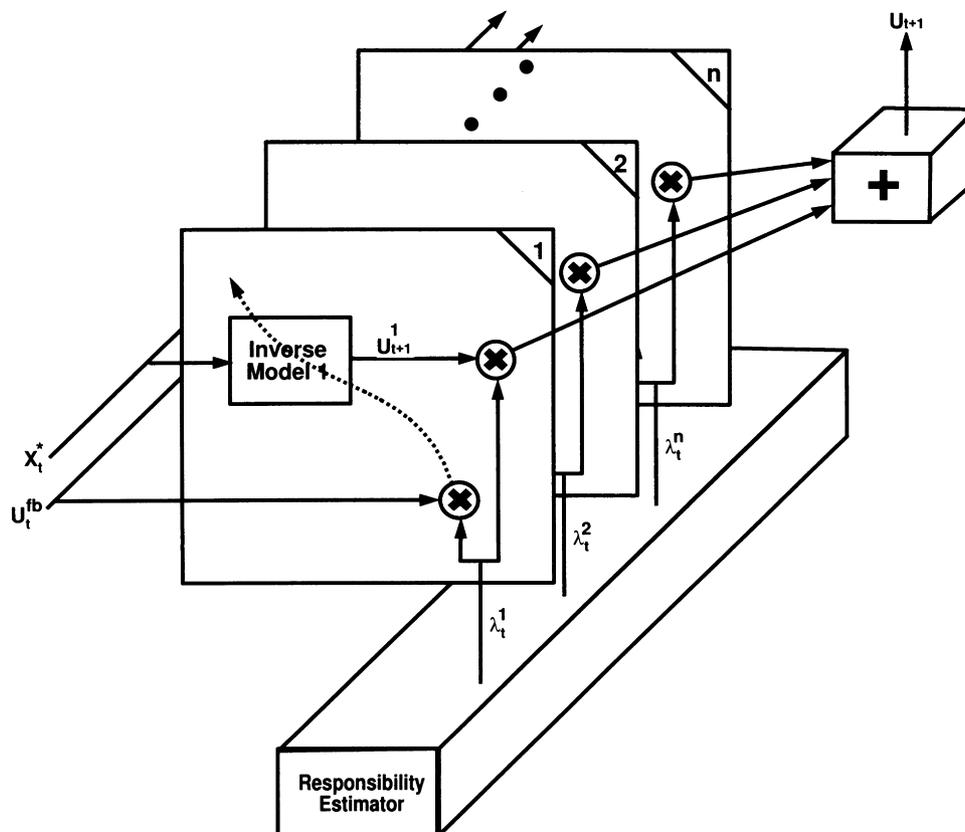


Fig. 2. Multiple inverse models for control. The inverse model component of the full model is shown. Each inverse model produces a feedforward motor command and receives as error the feedback motor command (see Section 6) weighted by the responsibility estimate of its paired forward model. This ensures that the inverse model learns the appropriate control for the context under which its paired forward model makes accurate predictions. The responsibility signals also weight the contribution of each inverse modules' output to the final feedforward motor command. The dotted lines passing through the models is the training signal for learning.

state. The total motor command generated by the whole set of  $n$  inverse models is given as the summation of outputs from these inverse models

$$u_t = \sum_{i=1}^n \lambda_t^i u_t^i = \sum_{i=1}^n \lambda_t^i \psi(\alpha_t^i, x_{t+1}^*) \quad (10)$$

once again using the responsibilities,  $\lambda_t^i$ , to weight the contributions. Again the responsibilities would be used to weight the learning of each inverse model (here we assume the desired motor command  $u_t^*$  is known but in a later section we show how this assumption can be relaxed)

$$\Delta \alpha_t^i = \epsilon \lambda_t^i \frac{d\psi_i}{d\alpha_t^i}(u_t^* - u_t^i) = \epsilon \frac{du_t^i}{d\alpha_t^i} \lambda_t^i (u_t^* - u_t^i) \quad (11)$$

Therefore, the responsibility signals are used in three ways—first to gate the learning of the forward models (Eq. (8)), second to gate the learning of the inverse models (Eq. (11)), and third to gate the contribution of the inverse models to the final motor command (Eq. (10)). In summary, each forward model receives the total motor command, and each models' prediction is compared with the true outcome. Only those forward models with small errors should adapt and those with large errors should learn little. This is mediated through the responsibilities  $\lambda_t^i$ . Conceptually speaking, if one forward model's prediction is good, its corresponding inverse model receives the major part of the motor error signal and its output contributes significantly to the final motor command. On the other hand, if the forward model's prediction is poor, its corresponding inverse model does not receive the full error and its output contributes less.

### 5.3. Multiple responsibility predictors—sensory contextual cues

While the system described so far can learn multiple controllers and switch between them based on prediction errors, it cannot provide switching before a motor command has been generated and the consequences of this action evaluated. To allow the system to switch controllers based on purely contextual information, cued by sensory signals, e.g. a tone or color, or endogenously generated by long term behavioral plans, we propose a third model, the responsibility predictor (RP). The input to this module,  $y_t$ , contains sensory information and cognitive plans (Fig. 3). Each RP produces a prediction of its modules responsibility

$$\hat{\lambda}_t^i = \eta(\gamma_t^i, y_t) \quad (12)$$

where  $\gamma_t^i$  are the parameter of a function approximator  $\eta$  (e.g. neural network weights). These estimated responsibilities can then be compared to the actual responsibilities generated from the responsibility estimator. These error signals can be used to update the weights of the RP by supervised learning.

For example, objects which look metallic are usually

heavier than wooden objects. Therefore, when picking up metal objects the RP would learn to predict a high responsibility for the module suitable for picking up heavy objects based on the sensory cues which determine whether an object looks metallic. Eventually, the RP would produce this high responsibility estimate based purely on the sensory signals before the feedback mechanism could make the actual responsibility of the module high.

Finally, a mechanism is required whereby the responsibility estimates emanating from the responsibility estimator are determined by both the responsibility estimate of the feedforward RP and feedback forward model. We determine the final value of responsibility by multiplying the transformed feedback errors by the feedforward responsibility  $\hat{\lambda}_t^i e^{-|x_t - \hat{x}_t^i|^2/\sigma^2}$  and then normalizing across the modules within the responsibility estimator (Fig. 3).

### 5.4. Probabilistic interpretation of responsibility estimation

The multiple paired forward–inverse model has a natural probabilistic interpretation. The estimates of the responsibilities produced by the RP can be considered as prior probabilities because they are computed before the movement execution based only on extrinsic signals and do not rely on knowing the consequences of the action. Once an action takes place, the responsibility calculated from the forward models' errors can be determined and this can be thought of as the likelihood after the movement execution based on knowledge of the result of the movement. The final responsibility which is the product of the prior and likelihood, normalized across the modules, represents the posterior. Adaptation of the RP ensures that the prior probability becomes closer to the posterior probability.

## 6. Physiologically plausible learning

The model described above assumes that the desired motor command is available to train multiple inverse models. This assumption is implausible for biological motor learning, and a more sophisticated physiological computational model involves using the feedback-error-learning model (Kawato et al., 1987; Kawato and Gomi, 1992) with the above simple model.

$$u_t = u_t^{\text{ff}} + u_t^{\text{fb}} \quad (13)$$

$$u_t^{\text{fb}} = g(x_t^* - x_t) \quad (14)$$

$$\Delta \alpha_t^i = \epsilon \lambda_t^i \frac{d\psi_i}{d\alpha_t^i} u_t^{\text{fb}} = \epsilon \frac{du_t^i}{d\alpha_t^i} \lambda_t^i u_t^{\text{fb}} \quad (15)$$

The total motor command fed to the motor apparatus is the summation of the total feed-forward motor command  $u_t^{\text{ff}}$  and the feedback motor command  $u_t^{\text{fb}}$ . The feedback motor command can be calculated from the difference between the desired movement pattern  $x_t^*$  and the actual movement

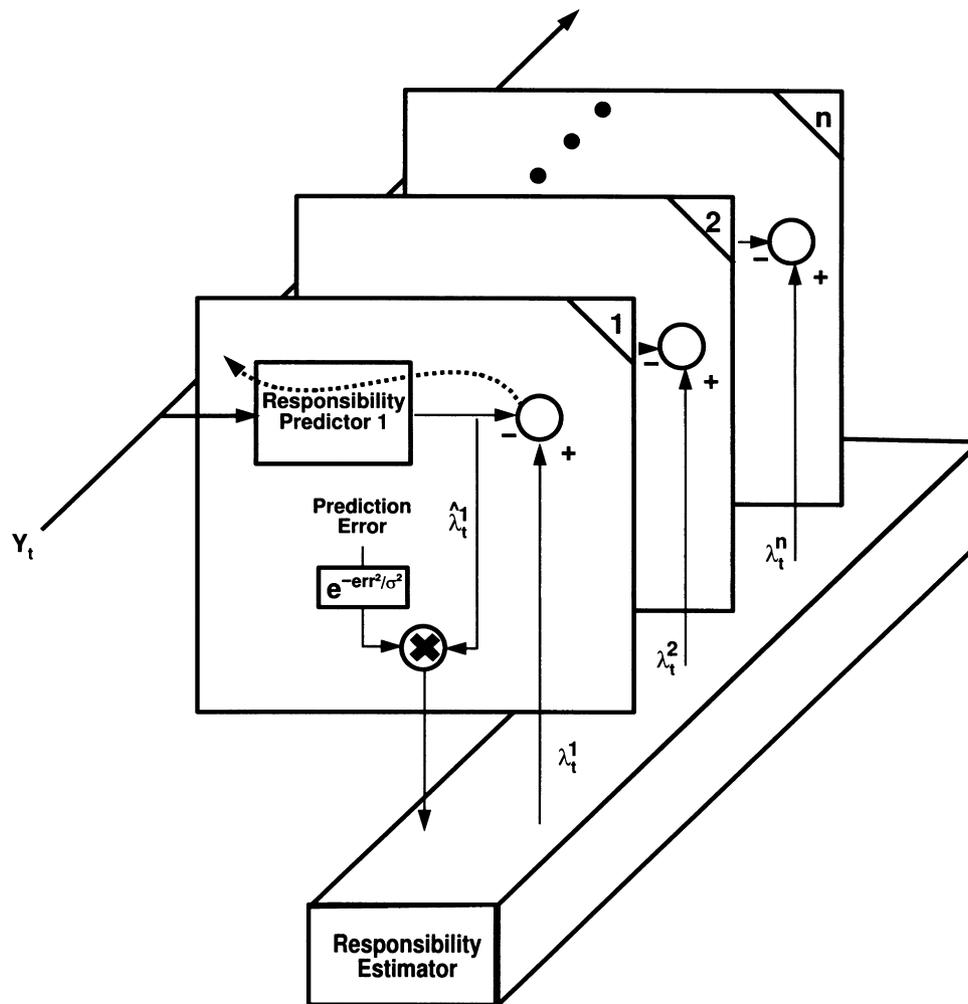


Fig. 3. Multiple responsibility predictors for sensory switching. The responsibility predictor (RP) component of the full model is shown. Each RP model produces an estimate of the module's responsibility based on the sensory input. The responsibility estimate is used as a training signal for the RP model. The estimate from the RP is multiplied by the transformed error from its paired forward model. The responsibility estimator normalizes these values to produce the responsibility for the module based on the feedforward RP and feedback forward model signals. The dotted lines passing through the models is the training signal for learning.

pattern  $x_t$  through an appropriate function  $g$  (usually a PID controller or PAD controller is assumed appropriate). In the learning (Eq. (15)), the feedback motor command is used as the error signal for the motor command.

A schematic of the inputs and outputs and internal structure within a module is shown in Fig. 4.

## 7. Comparison with other modular systems

We now compare the current model with the previous computational model in which multiple modules can be acquired through learning. Narendra et al. (1995) and Narendra and Balakrishnan (1997) have proposed multiple models, each of which describes a different environment, with switching for control. There are several major differences between Narendra's framework and ours, although they seem quite similar at first sight. In their approach, the

identification errors for forward models were used to determine switching and the controller was chosen for which the identification error was smallest. Therefore, only one controller could be active at any given time as opposed to the blending approach we have chosen and which we believe is a fundamental component of skilled motor control. Their model had no learning of the controllers and little in the forward model, and can only switch based on execution and not based on purely sensory cues. In contrast, our multiple paired model can learn both internal models from a naive state and can switch control based on purely sensory cues, mediated by the RP. These differences come from essential differences in backgrounds and research objectives. The former architecture is intended to be of practical use in control, and the latter for biological motor learning and control. The latter starts from scratch, which is out of the question with regard to stability, safety and other engineering requirements in the former approach. So, the

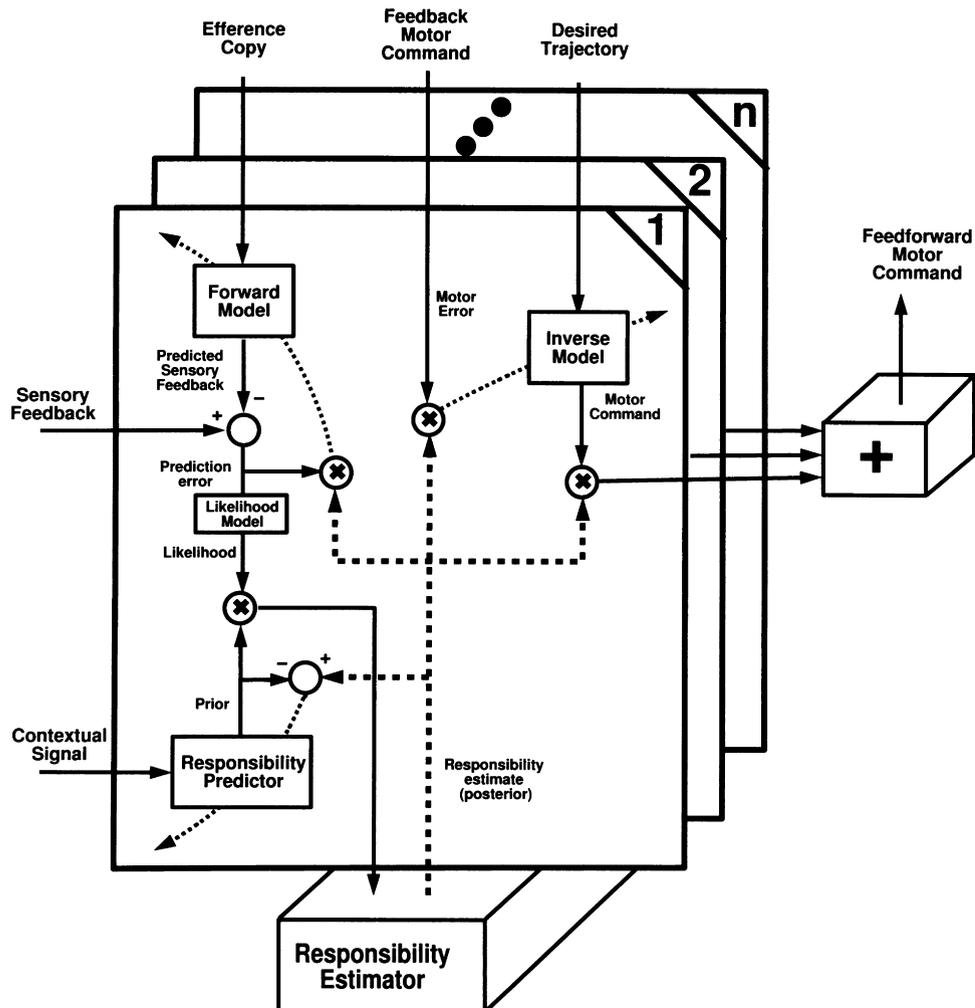


Fig. 4. A single module within the multiple paired internal model. The thick dashed line shows the central role of the responsibility estimators' signals. Dotted lines passing through models are training signals for learning. The exponential transform of the errors has been replaced by a more general likelihood model.

existence of a set of appropriate controllers is assumed, and blending outputs from a number of controllers is not considered. The former has much utility in current application in engineering but not in biology. The latter is a biological model and should show much utility in future applications.

Based on the principle of divide-and-conquer, a general computational strategy for designing modular learning systems is to treat the problem as one of combining multiple models, each of which is defined over a local region of the input space. Such a strategy has been introduced in the "mixture of experts" architecture for supervised learning (Jacobs et al., 1991b; Jordan and Jacobs, 1994; Cacciatore and Nowlan, 1994). The architecture involves a set of function approximators known as expert networks or modules (usually neural networks) that are combined by a classifier known as a gating network. These networks are trained simultaneously so as to split the input space into regions where particular experts can specialize. The gating network uses soft splits of the input data, thereby allowing data to be

processed by multiple experts. The contribution of each is modulated by the gating module's estimate of the probability that each expert is the appropriate one to use. Each expert is assumed to be responsible for a Gaussian region of input space which leads to the gating unit using a multinomial logit model to partition the input space. This model has been proposed for high-level vision (Jacobs et al., 1991a) and for the role of the basal ganglia during sensorimotor learning (Graybiel et al., 1994). The mixture of experts approach has been extended to a recursively-defined hierarchical mixture of experts (HME) architecture in which a tree of gating networks combines the expert networks into successively larger groupings that are defined over nested regions of the input space (Jordan and Jacobs, 1992). A maximum likelihood learning algorithm for the HME architecture has been derived (Jordan and Jacobs, 1994) based on the Expectation-Maximization (EM) principle from statistics (Dempster et al., 1977). The multiple experts model has been extended to deal with unsupervised learning in which the experts learn to partition time

series into components, with each component captured by an expert (Pawelzik et al., 1996; Ghahramani and Hinton, 1998).

Gomi and Kawato (1993) combined the feedback-error-learning approach and the mixture-of-experts modular architecture to learn multiple inverse models for multiple manipulated objects. They used both the visual signal of manipulated objects and internal signals, e.g. somatosensory feedback and efference copy of the motor command. It proved quite difficult to acquire multiple inverse models using this architecture and many parameters had to be carefully chosen. On the other hand, using the multiple paired forward–inverse model approach, the same task is easily solved (Haruno, Wolpert and Kawato, personal communication).

The essential difference between the mixture-of-experts architecture and ours is whether the one gigantic central processor or a set of parallelly-distributed small processors are used for calculating responsibility signals or “gating” signals. We believe that for any real world problem, calculating responsibility signals by a single gating network amounts to dividing very high dimensional state space with highly non-linear and complicated boundaries. This is a formidable and hopeless task. The other essential difference is the distinction between the inputs to the gating network and the inputs to our forward model and the RP. Although the gating network receives both the contextual and intrinsic signals to compute the gating signal, our forward model receives only intrinsic signals and the RP receives only the contextual signal. Thus, the notion of prior and likelihood probability is more physical and concrete. The prior is before the movement execution and the likelihood is after the movement execution.

One of the major reasons why the two approaches are very different comes from the objectives of the architecture. Although the mixture-of-experts architecture is intended to be very general as a universal tool for a function approximation, ours is intended to be a scheme for sensory-motor control. In general function approximation, it does not make sense to divide the input variables into contextual and internal ones. Moreover, using forward functions and their inverses does not generally help to solve the problem.

## 8. Model predictions

In this section, we will consider the specific, but as yet untested, predictions of the multiple paired forward–inverse model. We first focus on possible psychophysical experiments before briefly discussing the neurophysiological predictions.

### 8.1. Separation of forward and inverse models

The multiple paired forward–inverse model is based on a fundamental separation of the mappings into inverse and

forward models. One prediction is that the forward models have a primary role in motor learning. Therefore, during motor learning, the forward model will learn to predict before the inverse model learns to control. This could be tested in a number of ways. While the accuracy of the inverse model can be assessed by examining the change in performance errors during learning of a novel task, testing the forward model requires one of a number of indirect approaches. The first is to use a system which is known to require predictive forward models for its achievement. Such a system is seen in grip-force modulation, in which grip force is modulated, in an anticipatory fashion to the expected load force generated by the hand (Johansson and Westling, 1984). When the object is changed in a novel way, such modulation has to be re-learned (Flanagan and Wing, 1997) and this time-course can be used to estimate the changes in performance of the forward model. Alternatively, the predictions of the consequences of a motor command on the hand can be assessed by examining coordinative behavior. For example, when the eye tracks the hand in the dark it does so with no lag, showing that the eye is able to anticipate or predict the motion of the hand. It would be expected that under a perturbation to the hand, the eye would learn to re-follow the hand only when the new forward dynamics had been learned. A comparison of the learning rates of the forward and inverse model could therefore be used to test the primacy of the forward models.

### 8.2. Switching depends on prediction rather than control error

A fundamental aspect of the model is that switching should be based on prediction errors rather than performance errors. Experimentally, the two errors can be dissociated using altered visual feedback paradigms (Wolpert et al., 1995a). This allows control of the relationship between a target location and the perceived hand location (performance error), and the relationship between the perceived hand and actual hand location (which will affect the prediction error) to be dissociated. By maintaining the prediction error while the actual context changes, it should be possible to delay switching and new learning appropriately. Similarly, by maintaining the performance error but altering the prediction error, switching to an inappropriate control strategy should be seen.

### 8.3. Learning multiple modules

The model also predicts that it should be possible to learn multiple behavioral tasks and that the further apart the contexts are the easier it should be to learn them. For example, it has been shown that after learning one task, learning a second one can overwrite the first—this known as retrograde interference (Brashers-Krug et al., 1996). However, in these studies no context was given to the movement, and

the context of interacting with a robot may have caused the same module to be used for both tasks. This interference decreased as the duration between the learning phases was increased. Allowing time to pass may allow the context to change, as time is likely to play a factor in the context, thereby allowing new learning. The prediction of the multiple paired forward–inverse model is that if the contexts of the movements were made sufficiently different then the fields should not interfere with each other. Although it has been shown that simple lights do not perform such a function (Gandolfo et al., 1996), these are perhaps too behaviorally insignificant to cause separate learning. Learning to pick up objects with novel dynamics, in which the objects are indistinguishable or highly different, could be used to examine these possibilities.

#### 8.4. *Mixing multiple modules*

Another important prediction is that having learned a set of modules, their outputs should be able to be appropriately combined based on prior or posterior cues. This could be achieved by having subjects make movements in multiple simple force fields generated by a force feedback system. For example, having learned to move on two fields A and B, the model predicts that subjects should perform better on linear combinations of fields A and B compared to fields which have similar complexity but are unrelated to A and B. This could be used to test whether the motor learning system is capable of combining compact representations into more complex behaviors.

#### 8.5. *Decomposition of learning*

One prediction of the model is that the CNS should be able to extract compact representations of motor primitives during a learning task. Subjects could be exposed to different visuomotor perturbations which, unknown to them, will be constructed from a set of primitives. For example, the presented perturbations might be combinations of three primitives A, B and C. Subjects will be exposed to combined perturbations such as B, C, A + B, B + C, A + C, A + B + C, and the null perturbation, but never A in isolation. An efficient coding of these perturbations would be to represent the three primitives and switch them on or off appropriately. This hypothesis could be assessed by examining whether A has been learned having excluded its isolated presentation from the learning process.

#### 8.6. *Neurophysiological predictions*

In principle, multiple paired forward and inverse models could be located anywhere in the brain. However, many lines of investigation, both theoretical and experimental, suggest that the cerebellum is a very promising candidate (Ito, 1970; Kawato et al., 1987; Miall et al., 1993; Kawato and Gomi, 1992; Kawato and Gomi, 1993; Miall and

Wolpert, 1996). Wolpert et al. (1998) in a recent review summarized behavioral, anatomical and physiological data which directly and indirectly support the existence of both inverse models and forward models in the cerebellum.

Some imaging studies also indirectly support our hypothesis. For example, mental motor imagery is known to activate the cerebellum (Decety et al., 1990). Both forward models and inverse models are expected to be utilized in mental simulation of the movement. Forward models would be used, in place of the motor apparatus, to simulate the results of non-performed actions on an imaginary controlled object, and inverse models would be required to generate the motor command.

In a motor learning study using fMRI, Imamizu et al. (1997) demonstrated activation spots in the lateral posterior part of the cerebellum which persisted after learning to use a new tool was accomplished, even though the motor performance errors had returned to the levels of the baseline scanning periods. Although we cannot tell whether forward or inverse models are learned and reflected in these activation spots, the data certainly suggest that some kind of internal model of an external tool is acquired after motor learning in the cerebellum.

If multiple paired forward–inverse models are located in the cerebellum, we can at least make two predictions for imaging studies. The first involves subjects learning to use multiple tools or perform under multiple visuo-motor transformations. After learning, the cerebellum would be scanned under at least two different conditions using different tools or visuomotor transformations. The prediction is that two different areas, or sets of areas, of the cerebellum should be activated under these two different conditions.

A second prediction is that different areas of the cerebellum would be activated when either the prediction error or control error are independently manipulated. Such a study would use the methodology described in Section 8.2 to dissociate the control error and prediction error. During a motor task, e.g. tool use or a visuo-motor transformation, which activates some regions of the cerebellum, we predict that only a subset of these activated regions will be vigorously activated when the prediction error is artificially increased by experimental manipulation. This activated locus corresponds to the forward model. On the other hand, a different subset of the above activated regions should be more vigorously activated when the control error is artificially increased. This corresponds to the inverse model. We expect that these two subsets are close to each other. But, even if they were far apart, it would not contradict the computational model, although such an arrangement would require long range connections between the forward and inverse models.

We plan to map the computational circuit diagram of our model onto the neural networks in and around the cerebellum. We plan to fully develop this network model in another paper.

## 9. Conclusions

In conclusion, we have presented a new model based on multiple paired forward and inverse models, which is capable of motor learning and control in a modular network. The problem of selecting the appropriate modules is solved by generating a responsibility signal for each module based both on the consequences of performed actions, as estimated by the forward models, and on sensory signals, as estimated by the responsibility predictor. Within each module, the inverse and forward models are tightly coupled, by the responsibility signal, during motor learning. This architecture can, therefore, simultaneously learn the multiple inverse models necessary for control as well as how to select the inverse models appropriate for a given environment.

## Acknowledgements

We thank Zoubin Ghahramani for helpful discussions on the probabilistic interpretation of the model and Sarah Blakemore for comments on the manuscript. This work was supported by grants from the Wellcome Trust, the Medical Research Council, the Royal Society, the BBSRC and Human Frontier Science Project.

## References

- Baker J.F., Perlmutter S.I., Peterson B.W., Rude S.A., & Robinson F.R. (1987). Simultaneous opposing adaptive changes in cat vestibulo-ocular reflex directions for two body orientations. *Exp. Brain Res.*, *69*, 220–224.
- Brashers-Krug T., Shadmehr R., & Bizzi E. (1996). Consolidation in human motor memory. *Nature*, *382*, 252–255.
- Cacciatore, T.W., & Nowlan, S.J. (1994). Mixtures of controllers for jump linear and non-linear plants. In J.D. Cowan, G. Tesauro & J. Alsppector (Eds.), *Advances in neural information processing systems 6* (pp. 719–726). San Francisco, CA: Morgan Kaufman.
- Decety J., Sjöholm H., Ryding E., Sternberg G., & Ingvar D.H. (1990). The cerebellum participates in mental activity. *Brain Res.*, *553*, 313–317.
- Dempster A.P., Laird N. M., & Rubin D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc. B*, *39*, 1–38.
- Flanagan J.R., & Wing A.M. (1997). The role of internal models in motion planning and control: evidence from grip force adjustments during movements of hand-held loads. *J. Neurosci.*, *17*, 1519–1528.
- Gandolfo F., Mussa-Ivaldi F.A., & Bizzi E. (1996). Motor learning by field approximation. *Proc. Natl. Acad. Sci.*, *93*, 3843–3846.
- Ghahramani, Z. & Hinton, G.E. (1998). Switching state-space models. (Submitted).
- Ghahramani Z., & Wolpert D.M. (1997). Modular decomposition in visuomotor learning. *Nature*, *386*, 392–395.
- Gomi H., & Kawato M. (1993). Recognition of manipulated objects by motor learning with modular architecture networks. *Neural Networks*, *6*, 485–497.
- Graybiel A.M., Aosaki T., Flaherty A.W., & Kimura M. (1994). The basal ganglia and adaptive motor control. *Science*, *265* (5180), 1826–1831.
- Hay J.C., & Pick H.L. (1966). Gaze-contingent prism adaptation: Optical and motor factors. *J. Exp. Psychol.*, *72*, 640–648.
- Imamizu H., Miyauchi S., Sasaki Y., Takino R., Pütz B., & Kawato M. (1997). Separated modules for visuomotor control and learning in the cerebellum: a functional MRI study. *NeuroImage*, *5*, S598.
- Ito M. (1970). Neurophysiological aspects of the cerebellar motor control system. *Int. J. Neurol.*, *7*, 162–176.
- Ito, M. (1984). *The cerebellum and neural control*. New York: Raven Press.
- Jacobs R.A., Jordan M.I., & Barto A.G. (1991). Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. *Cognitive Science*, *15* (2), 219–250.
- Jacobs R.A., Jordan M.I., Nowlan S.J., & Hinton G.E. (1991). Adaptive mixture of local experts. *Neural Computation*, *3*, 79–87.
- Johansson R.S., & Westling G. (1984). Roles of glabrous skin receptors and sensorimotor memory in automatic-control of precision grip when lifting rougher or more slippery objects. *Exp. Brain Res.*, *56*, 550–564.
- Jordan, M.I. (1995). Computational aspects of motor control and motor learning. In H. Heuer & S. Keele (Eds.), *Handbook of perception and action: Motor skills*. New York: Academic Press.
- Jordan, M.I., & Jacobs, R.A. (1992). Hierarchies of adaptive experts. In J. Moody, S. Hanson, & R. Lippmann (Eds.), *Advances in neural information processing* (pp. 985–993). San Mateo, CA: Morgan Kaufmann.
- Jordan M.I., & Jacobs R.A. (1994). Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, *6*, 181–214.
- Jordan M.I., & Rumelhart D.E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science*, *16*, 307–354.
- Kawato, M. (1990). Feedback-error-learning neural network for supervised learning. In R. Eckmiller (Ed.), *Advanced neural computers* (pp. 365–372). Amsterdam: North-Holland.
- Kawato M., Furawaka K., & Suzuki R. (1987). A hierarchical neural network model for the control and learning of voluntary movements. *Biol. Cybern.*, *56*, 1–17.
- Kawato M., & Gomi H. (1992). The cerebellum and VOR/OKR learning models. *Trends Neurosci.*, *15*, 445–453.
- Kawato M., & Gomi H. (1993). Theories of motor learning by the cerebellum—reply. *Trends Neurosci.*, *5*, 177–178.
- Kohler, I. (1950). Development and alterations of the perceptual world: conditioned sensations. *Proceedings of the Austrian Academy of Sciences*, p. 227.
- Kravitz J.H. (1972). Conditioned adaptation to prismatic displacement. *Perception and Psychophysics*, *11*, 38–42.
- Kravitz J.H., & Yaffe F. (1972). Conditioned adaptation to prismatic displacement with a tone as the conditional stimulus. *Perception and Psychophysics*, *12*, 305–308.
- Kuperstein M. (1988). Neural model of adaptive hand-eye coordination for single postures. *Science*, *239*, 1308–1311.
- Martin T.A., Keating J.G., Goodkin H.P., Bastian A.J., & Thach W.T. (1996). Throwing while looking through prisms. II. Specificity and storage of multiple gaze throw calibrations. *Brain*, *119*, 1199–1211.
- McGonigle B.O., & Flook J.P. (1978). Long-term retention of single and multistate prismatic adaptation by humans. *Nature*, *272*, 364–366.
- Miall R.C., Weir D.J., Wolpert D.M., & Stein J.F. (1993). Is the cerebellum a Smith predictor?. *J. Motor Behav.*, *25* (3), 203–216.
- Miall R.C., & Wolpert D.M. (1996). Forward models for physiological motor control. *Neural Networks*, *9* (8), 1265–1279.
- Miller W.T. (1987). Sensor-based control of robotic manipulators using a general learning algorithm. *IEEE J. Robotics and Automation*, *3*, 157–165.
- Narendra K.S., & Balakrishnan J. (1997). Adaptive control using multiple models. *IEEE Trans Automatic Control*, *42* (2), 171–187.
- Narendra K.S., Balakrishnan J., & Ciliz M.K. (1995). Adaptation and learning using multiple models, switching, and tuning. *IEEE Control Systems Magazine*, *15* (3), 37–51.
- Pawelzik K., Kohlmorgen J., & Müller K.-R. (1996). Annealed competition of experts for a segmentation and classification of switching dynamics. *Neural Computation*, *8*, 340–356.
- Shadmehr R., & Mussa-Ivaldi F. (1994). Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.*, *14* (5), 3208–3224.
- Shelhamer, M., Robinson, D.A., & Tan, H.S. (1991). Context-specific gain

- switching in the human vestibuloocular reflex. In B. Cohen, D.L. Tomko, & F. Guedry (Eds.), *Annals of the New York Academy of Sciences* (Vol. 656, pp. 889–891). New York: New York Academy of Sciences.
- Sutton R.S., & Barto A.G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. *Psychol. Rev.*, 88, 135–170.
- Welch R.B. (1971). Discriminative conditioning of prism adaptation. *Perception and Psychophysics*, 10, 90–92.
- Welch, R.B. (1986). Adaptation of space perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1:24). New York: Wiley.
- Welch R.B., Bridgeman B., Anand S., & Browman K.E. (1993). Alternating prism exposure causes dual adaptation and generalization to a novel displacement. *Perception and Psychophysics*, 54 (2), 195–204.
- Wolpert D.M. (1997). Computational approaches to motor control. *Trends Cogn. Sci.*, 1 (6), 209–216.
- Wolpert D.M., Ghahramani Z., & Jordan M.I. (1995a). Are arm trajectories planned in kinematic or dynamic coordinates? An adaptation study. *Exp. Brain Res.*, 103 (3), 460–470.
- Wolpert D.M., Ghahramani Z., & Jordan M.I. (1995b). An internal model for sensorimotor integration. *Science*, 269, 1880–1882.
- Wolpert, D.M., Miall, R.C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends Cogn. Sci.*