In P.G. Morasso and V. Sanguineti (eds.) *Self-Organization, Computational Maps and Motor Control,* 117–147. Amsterdam: North–Holland, 1997.

Computational Models of Sensorimotor Integration

Zoubin Ghahramani[*]
Daniel M. Wolpert[†]
Michael I. Jordan[‡]

Abstract

The sensorimotor integration system can be viewed as an observer attempting to estimate its own state and the state of the environment by integrating multiple sources of information. We describe a computational framework capturing this notion, and some specific models of integration and adaptation that result from it. Psychophysical results from two sensorimotor systems, subserving the integration and adaptation of visuo-auditory maps, and estimation of the state of the hand during arm movements, are presented and analyzed within this framework. These results suggest that: (1) Spatial information from visual and auditory systems is integrated so as to reduce the variance in localization. (2) The effects of a remapping in the relation between visual and auditory space can be predicted from a simple learning rule. (3) The temporal propagation of errors in estimating the hand's state is captured by a linear dynamic observer, providing evidence for the existence of an internal model which simulates the dynamic behavior of the arm.

1 Introduction

All higher organisms are able to integrate information from multiple sensory modalities and use this information to select and guide movements. At the outset, this problem seems formidable. Information

[*]Department of Computer Science, University of Toronto, Toronto, ON M5S 1A4, Canada. [†]Sobell Department of Neurophysiology, Institute of Neurology, Queen Square, London WC1N 3BG, United Kingdom. [‡]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

arriving into each sense codes for quite different aspects of the environment: Audition senses changes in pressure on the eardrum; vision detects photons on the retina; the sense of smell recognizes molecules in the olfactory bulb. The central nervous system accomplishes the task of extracting the commonalities in this information, and integrating these into unified percepts. This seemless integration of information not only underlies perception but also the production of movement. A single reaching movement, for example, may require convergence of information from the visual, proprioceptive, and motor systems.

The goal of this chapter is to outline a computational theory of sensorimotor integration. While each sensory modality and motor subsystem is distinct in its functioning, there are common elements to the problem of integrating multiple sources of information which can be captured within a computational framework. As in other areas of neuroscience, we appeal to the formal analyses of this problem that have been made within statistics, computer science, and engineering. Thus, in the tradition of Marr (1982), we seek to understand sensorimotor integration by asking: (1) what is the problem from a computational point of view, (2) how can this problem be solved, and (3) how would such a solution be implemented in the brain.

Of course, no theory is useful if it does not make predictions; one advantage of computational theories is that they often make quite precise quantitative predictions. After we have outlined several models of sensorimotor integration, we present recent data which allows us to assess these models critically. These data were obtained from behavioral experiments dealing with (1) the multisensory integration system which spatially localizes visual and auditory targets, and (2) the sensorimotor integration system which estimates the location of the hand during arm movements. By examining the behavioral data in light of the model predictions we narrow the search for theories of visuo-auditory integration and adaptation, and posit the existence of an internal model for sensorimotor integration.

1.1 The Need for Integration

While it may be clear that the central nervous system (CNS) needs to integrate information from different senses, it is nonetheless useful to examine the possible specific advantages this integration may provide. The study of robotics suggests the following advantages may be gained by a system which combines multiple information sources (Durrant-Whyte 1988; Abidi & Gonzalez 1992):

- Multiple sensors provide redundancy, which can both reduce the overall uncertainty of sensory estimates and increase the reliability in the case of sensor failure.
- Complementary information may be gained from the different senses. By integrating information across sensors, it may be possible to derive information that is impossible to derive using each individual sensor (e.g. stereo vision is only possible by using information from both eyes).
- More timely information may be obtained through parallelism, as each sensor may have a different latency. For example, the early stages of visual information processing can take around 150 ms, as compared to 30 ms for auditory information. Such differences in latency may be traded-off with differences in accuracy in order to obtain a rapid but crude sensory estimate early on, which is later refined by inputs from other sensors.

By translating the intuitive notion of the advantage gained from integration into a quantitative measure, or *cost function*, it becomes possible to formulate a computational theory of sensorimotor integration. For example, the above sources of advantage could be quantified through costs based on "uncertainty in sensor estimate", "probability of failure", or "latency of response". Such a theory is useful both for design and modeling purposes. Given a cost function, one can define what is meant by an *optimal* integration of several information sources, an approach that is commonplace in engineering. To understand the central nervous system we will make use of a reverse-engineering approach: We use the behavior of the system to infer a cost function whose minimization would reproduce this behavior. In this regard our approach is very similar to the optimization framework that has been used extensively in the study of movement planning (Nelson 1983, Hogan 1984, Flash & Hogan 1985, Uno et al. 1989, Wolpert et al. 1995a).

Although all the above uses of integration may play an important role for the organism, we will focus on only one: reducing the uncertainty in sensor estimates. We view the perceptuomotor system from the point of view of an observer attempting to estimate some relevant attribute of the environment such as the location of a target (Gibson 1961, Richards 1988, Bennett et al. 1989). We test the hypothesis that the observer is integrating multiple information sources so as to minimize the uncertainty in this estimate. There are several ways in which this cost can be defined, which we will explore in section 2. We also explore the relation between sensorimotor integration and adaptation. Given a particular cost for integration, one can derive a learning rule for adaptation consistent with that cost. We expand on this in section 2.3 and report on some experiments in section 3.

The observer approach has been often used in the study of purely perceptual systems (e.g. Nakayama & Shimojo 1992). One difference between perceptual and sensorimotor systems is that, in the latter, the observer may also need to dynamically integrate reafferent sensory signals and copies of motor efference that arise during movement (Wolpert, Ghahramani & Jordan 1995b). We explore this from a computational perspective in section 2.2.1 and report on one relevant experiment in section 4.

2 The Computational Model

The presence of information common to multiple sensory modalities poses two challenging computational problems for the CNS. First, the signals from different modalities must be converted into a common representation appropriate for fusion. Second, using some sensible combination rule, signals in this common representation must be fused.

Although these two problems need not be solved sequentially, or by separate neural processes, the distinction appears to be useful from a computational perspective. Furthermore, the existence of multiple aligned sensory maps in sensorimotor areas such as the superior colliculus suggests that this distinction is also relevant at the neural level (Wickelgren 1971, Harris et al. 1980, Knudsen & Knudsen 1989a, Stein & Meredith 1993). Our focus will be on the latter problem, which we refer to as the *integration problem*, although we will also discuss briefly the former problem, which we refer to as the *coordinate transformation problem.*

2.1 The Coordinate Transformation Problem

Consider a system which receives inputs from two sources, X and Y, which could correspond for example to two sensory modalities. In order to transform these sources into a common representation, the system must first filter information that is common to both modalities, while rejecting that which is not. For example, the location of activity on the retina and an auditory interaural time difference both reflect spatial attributes of a visuo-auditory stimulus. In this case, the system would need to extract this commonality and suppress other attributes, such as color and pitch, in order to generate a map registering both visual and auditory space. While it is plausible that the separation of these sources may be largely driven by innate wiring of the CNS, we will ask to what extent a computational theory based on activity-dependent changes could account for it.

The idea of extracting common information from different sensory modalities can be phrased succinctly in the language of information theory. Information is defined as the capacity for a signal to reduce a system's uncertainty (Cover & Thomas 1991). The information content of a source X is defined as (Shannon 1948):

$$H(X) = -\sum_{j=1}^{n} P(X = x_j) \log P(X = x_j), \tag{1}$$

where $P(X = x_j)$ is the probability of receiving input x_j. (For continuous signals, a limiting argument is used to convert this sum into an integral.) The information common to two transformed signals $f(X)$ and $g(Y)$, known as the *mutual information*, is defined as:

$$I(f(X), g(Y)) = H(f(X)) + H(g(Y)) - H(f(X), g(Y)). \qquad (2)$$

Thus, a natural goal for a multisensory system with two coordinate transformations f and g is to maximize $I(f(X), g(Y))$.

Building on a large literature on information-maximizing models of perceptual processing (Attneave 1954, Barlow 1961, Linsker 1986), Becker & Hinton (1992) proposed utilizing mutual information as the basis for an optimization algorithm that extracts information from multiple input streams. They showed that a model based on maximizing mutual information could discover stereo disparity from a random-dot stereogram, capturing interesting structure that is not present in any single input source.

Unfortunately, the idea of maximizing mutual information cannot capture one of the fundamental properties of coordinate transformations in the CNS: topographic organization. Any one-to-one transformation of f or g will not affect $I(f(X), g(Y))$, while potentially making the coordinate transformation between $f(X)$ and $g(Y)$ arbitrarily complex. Fortunately, it is possible to augment the mutual information cost function with a term incorporating topographic order (Ghahramani 1995; Chapter 5). Simulations indicate that using this augmented cost function, two mutually-aligned topographic maps can arise through activity-dependent learning. This suggests that the combination of information-theoretic principles with topographic organization may provide a basis for solving the coordinate transformation problem. In the remainder of this chapter, we will focus on the problem of integrating signals once they have been transformed into a common coordinate frame.

2.2 The Integration Problem

Consider n signals originating from separate sources which have already been converted into a common representation. The simplest observer operates under the assumption that each of these signals is a noisy measurement of some underlying quantity that is to be estimated, such as the location or motion vector of an object. The measurements x_i, $i = \{1, \ldots, n\}$, can be modeled by assuming that the underlying quantity x has been corrupted by adding noise ϵ_i:

$$x_i = x + \epsilon_i. \tag{3}$$

Which estimate of x is optimal depends on the cost function used. The statistical theory of maximum likelihood estimation suggests using as a cost the probability of the measurements given the estimate:

$$P(x_1, x_2, \ldots, x_n | x). \tag{4}$$

Assuming, for now, that each of the noise processes ϵ_i is independent, the likelihood can be factored:

$$P(x_1, x_2, \ldots, x_n | x) = \prod_{i=1}^{n} P(x_i | x). \tag{5}$$

This expression makes it clear that to obtain a maximum likelihood estimate (MLE) of x, the system must have a statistical model of the process generating the data $P(x_i|x)$. If each noise source has a zero-mean Gaussian distribution of differing variance σ_i^2, the MLE of x is given by

$$\hat{x} = \sum_{i=1}^{n} \frac{\sigma_i^{-2} x_i}{\sum_{j=1}^{n} \sigma_j^{-2}} = \sum_{i=1}^{n} w_i x_i, \tag{6}$$

where $w_i = \sigma_i^{-2} / \left(\sum_{j=1}^{n} \sigma_j^{-2} \right)$. This integration rule states that the optimal estimate linearly combines the signals, weighted by their inverse variances.

Integration rule (6) can also be obtained if we assume that all we know about each signal is its variance or uncertainty, and we wish to combine them linearly so as to minimize the variance of our estimate. The variance of this estimate is

$$\sigma_{\hat{x}}^2 = (\sum_{i=1}^{n} \sigma_i^{-2})^{-1}, \tag{7}$$

which is smaller than the variance of each of the signals and of any other unbiased estimator.

Finally, (6) can also be motivated from an information-theoretic framework by noting that the information content of a Gaussian is inversely related to its variance. Equation (6) therefore defines the unbiased linear estimate with maximal information content under a Gaussian noise model.

Keeping in mind these alternative interpretations, we refer to the estimate given by (6) as the *minimum variance estimate* (MVE). Extensions to non-independent noise, multivariate measurements, and other distributions can be readily obtained. We now focus on an extension of the MVE that is particularly relevant to sensorimotor integration.

2.2.1 The Kalman filter. A particularly useful and general form of estimator resulting from the minimum variance integration principle is the Kalman filter (Kalman & Bucy 1961). This extends the framework we have described in two ways. First, the value we wish to estimate, known as the *state*, is not constant in time but depends on the previous state through a linear dynamical equation:

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + \mathbf{w}_t, \tag{8}$$

where $\mathbf{u}_t$ is some input or control signal that the system can observe at time t, and $\mathbf{w}_t$ is zero mean noise. Second, the measurements observed, denoted by $\mathbf{y}_t$, are related to the state through another linear equation:

$$\mathbf{y}_t = C\mathbf{x}_t + \mathbf{v}_t, \tag{9}$$

where $\mathbf{v}_t$ is again zero mean noise. The basic idea of the Kalman filter is that an optimal estimate of the state, $\hat{\mathbf{x}}_{t+1}$, can be obtained by fusing the input $\mathbf{u}_t$, the observations $\mathbf{y}_t$, and the previous state estimate $\hat{\mathbf{x}}_t$ using a model of the dynamical system. Based solely on the previous state, that is, before having observed $\mathbf{y}_t$, the best estimate of $\hat{\mathbf{x}}_{t+1}$ is clearly given by $A\hat{\mathbf{x}}_t + B\mathbf{u}_t$. Upon observing $\mathbf{y}_t$ this estimate is corrected via a term proportional to the error in the predicted observation, resulting in the following update rule:

$$\hat{\mathbf{x}}_{t+1} = A\hat{\mathbf{x}}_t + B\mathbf{u}_t + K_t[\mathbf{y}_t - C\hat{\mathbf{x}}_t]. \tag{10}$$

The matrix K_t is the *Kalman gain*, which weights the previous state estimate and the new input in proportion to their inverse variances.

The optimality of Kalman filters can also be stated in several ways. If the noise is Gaussian, the filter provides the maximum likelihood estimator in the sense previously described. However, if the noise is not Gaussian, the Kalman filter still provides the minimum variance *linear* estimator for the state (Goodwin & Sin 1984).

From the point of view of neuroscience, an interesting aspect of the Kalman filter is that it incorporates an internal model of the dynamics of the system being modeled. Based on computational principles alone, it has been proposed that the CNS uses an internal model in motor planning, control and learning (Ito 1984, Kawato et al. 1987, Jordan & Rumelhart 1992, Miall et al. 1993). Using a Kalman filter to model the propagation of state estimation errors during movement, it is possible to test empirical hypotheses concerning the existence and use of an internal model by the CNS. This is the topic of section 4.

2.3 From Integration to Adaptation

When the sensory inputs to an integration process are in disagreement, it is possible that one of them is miscalibrated. The optimal strategy for the nervous system in this case may involve adapting the interpretation of one of the sources, changing the relative weights of the sources, or both. Viewed in this way, the convergence of signals at the locus of integration provides a tool for recalibrating each of the sensory inputs. Thus, it would seem that the mechanisms underlying integration should be closely related to those underlying intersensory adaptation.

The goal of this section is to make explicit the connection between integration and adaptation by describing a method for deriving learning rules which are consistent with a particular integration rule. For example, in minimum variance integration the learning rule adapts each modality in proportion to the weighting of the other modalities. That is, for two modalities, the less dominant one will adapt more than the more dominant one. In the limit of complete adaptation, both modalities will converge to the minimum variance estimate.

Consider two signals, x_1 and x_2 with variances σ_1^2 and σ_2^2. The minimum variance estimator is given by

$$\hat{x} = w_1 x_1 + w_2 x_2,$$

where

$$w_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

and $w_2 = 1 - w_1$. If the two signals disagree, for example by a constant offset or bias, how much should each modality adapt to incorporate this bias? Perhaps the simplest supervised learning rule, known in various literatures as the delta rule, the Widrow-Hoff rule, or the LMS rule, and derivable as a maximum likelihood estimate under a Gaussian noise assumption, states that if a true target value is known, then each input should be adapted in the direction of this target (Widrow & Hoff 1960, Rumelhart & McClelland 1986, Hertz et al. 1991). Denoting the target value by x^*, and letting η be a small constant of proportionality—the *learning rate*—then the delta rule can be written

$$\Delta x_1 = \eta(x^* - x_1),$$

where Δx defines the change applied to x.

By assumption, the multisensory observer does not have access to an explicit teaching signal or true target—access to such a target would make perception trivial. However, by replacing the target with the minimum variance estimate of x we obtain the following interesting form of the delta rule:

$$\begin{aligned} \Delta x_1 &= \eta(\hat{x} - x_1) \\ &= \eta(w_1 x_1 + w_2 x_2 - x_1) \\ &= \eta w_2 (x_2 - x_1). \end{aligned} \tag{11}$$

We will call the learning rule given by (11) the *weighted delta rule* (WDR). It states that each modality should adapt in the direction of the other by an amount proportional to the weighting assigned to the *other* modality. For example, if the two modalities are vision and audition, then the WDR predicts that the auditory map should adapt more where the visual input is more dominant, where the visual dominance may be a function of spatial location or experimental conditions.

An alternative form of the weighted delta rule can be derived simply by stating that each modality adapts in proportion to how variable it is. This rule,

$$\Delta x_1 = \eta \sigma_1^2 (x_2 - x_1) \tag{12}$$

which we will call the *variance-weighted delta rule* (VWDR), can be derived from the maximum likelihood framework if each modality assumes that the other is its target.

It is easy to show that both the WDR and VWDR maintain the minimum variance estimate invariant over time, and converge with the mean estimate given by each modality equal to the minimum variance estimator (Ghahramani 1995). In the case of two modalities, the only difference between the WDR and the VWDR is that the normalization constant in the weights in the WDR has been absorbed into the learning rate of the VWDR. However, as will be shown later in this chapter, this difference can cause markedly differing predictions regarding the pattern of adaptation.

2.4 Other Models of Integration and Adaptation

2.4.1 Competitive integration. The principles presented so far could be termed *cooperative*, in the sense that an estimate is obtained by combining the contributions of all the sensory inputs. In contrast *competitive*, or *winner-take-all*, principles capture the notion that in the presence of disagreement, one of the senses may dominate and the others be ignored. Thus, for example, the competitive integration rule based on smallest variance can be stated as

$$\hat{x} = x_i \quad \text{iff} \quad \sigma_i^2 \leq \sigma_j^2 \quad \forall j. \tag{13}$$

As before, paralleling this integration rule is a competitive adaptation rule. Letting i index the dominant input (e.g. the input with the smallest variance) the learning rule can be written

$$\Delta x_j = \eta (x_i - x_j), \tag{14}$$

which is exactly the delta rule; the dominant modality acts as a target for the non-dominant ones. In the case of vision and audition, for example, if we assume that vision is dominant, the integration rule (13) predicts that in the presence of a visuo-auditory discrepancy complete visual capture will occur (e.g. the "ventriloquism" effect; Howard & Templeton 1966). Furthermore, (14) predicts that a persistent discrepancy will induce auditory adaptation, but no visual adaptation.

2.4.2 Stochastic integration. A different form of competitive integration occurs if the CNS selects between discrepant signals probabilistically. For example, simultaneous visual and auditory stimuli may cause a saccade to either of the two stimuli rather than to a location in between. This form of integration, which we will call *stochastic integration*, can also be based on a measure of variance or reliability. If the probability of choosing signal i is inversely proportional to its variance, $p_i \propto \sigma_i^{-2}$, we obtain

$$\hat{x} = \begin{cases} x_1 & \text{with prob. } p_1 \\ \vdots & \\ x_n & \text{with prob. } p_n. \end{cases} \tag{15}$$

Note that the probabilities, when normalized, are exactly equal to the weights $w_1, \ldots, w_n$ in the MVE, making this a stochastic version of the minimum variance estimator. The mean of this estimator is the MVE, however, its variance is guaranteed to be at least n times higher than the variance of the MVE. A testable prediction made by this rule is that the distribution of the estimates (i.e. responses) when two sensory modalities are stimulated will be bimodal, with the modes predictable from the responses to unisensory stimuli. The adaptation rule consistent with this integration rule uses the randomly selected signal as the target for the other signals. This has the interesting effect that all the modalities will also converge on the MVE.

3 Integration and Adaptation of Visual and Auditory Maps

The models proposed in the previous section make precise quantitative predictions both regarding how signals from several sensory modalities will be combined in order to produce a motor response, and the patterns of sensorimotor adaptation that will arise from an intersensory discrepancy. Using a psychophysical paradigm in humans, we have tested some of these predictions for the system involved in localizing visual and auditory targets (Ghahramani 1995, Ghahramani et al. 1995).

The basic experimental procedure consisted of measuring the biases (constant errors) and variances in localization of visual (V), auditory (A), and visuo-auditory (VA) stimuli. Subjects were presented with one of the three types of stimuli, randomly interleaved, and their goal was to point to the location of the stimulus as accurately as possible (Figure 1). Each of the models in the previous section predicts a different pattern of localization variances for the VA stimuli based on the subject's responses to the V and A stimuli separately.

The first observation to note is that visual localization is much less variable than auditory localization (Figure 2a). For both vision and audition, localization is best straight-ahead and increases in variability towards the periphery—a finding that is consistent with the existing literature (Mills 1958, Middlebrooks & Green 1991). The relative variances of visual and auditory localization suggest that vision provides much more reliable spatial information than audition throughout the azimuth. Indeed, when simultaneous visual and auditory stimuli are presented the variance in localization is not significantly different from the variance for visual stimuli alone (Figure 2a triangles). This finding is statistically consistent with the predictions of both the minimum variance integration rule (which would give vision a weighting of 0.9; Figure 2b), and the competitive integration rule (which would use only vision). This data is, however, inconsistent with the stochastic integration rule, which predicts that VA variance will be more than twice the visual variance.

a) b)

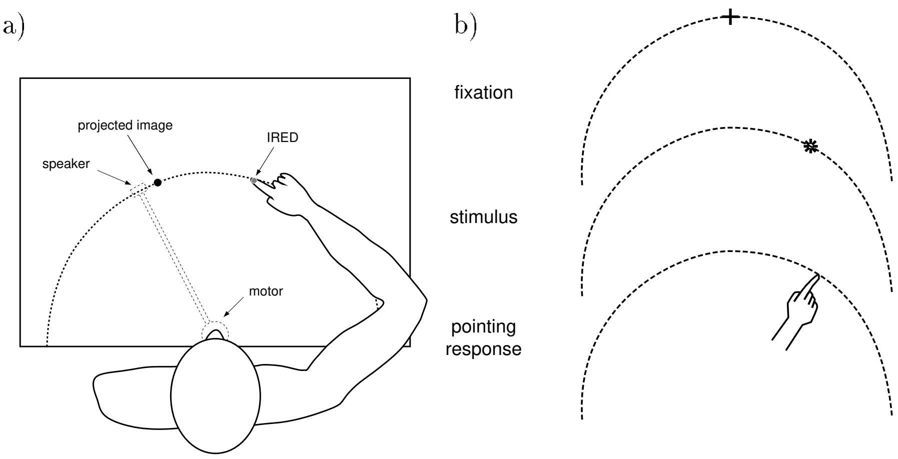

Figure 1: a) Experimental setup. Subjects are seated at a table with an Optotrak infrared marker mounted on their right index finger, which was used to record the pointing responses. Visual stimuli (5 mm white squares) were projected onto a screen on the table using an LCD projector. Auditory stimuli were presented using a small speaker (300-500 Hz, 75 dB tone at 20cm) directly below the screen, whose position was controlled by a stepper motor. b) Experimental paradigm. Trials started with fixation on the cross straight-ahead (0°). The cross disappeared and after 100 ms either a visual, auditory, or simultaneous visuo-auditory stimulus was presented for 100 ms. The subject then pointed to the perceived stimulus location.

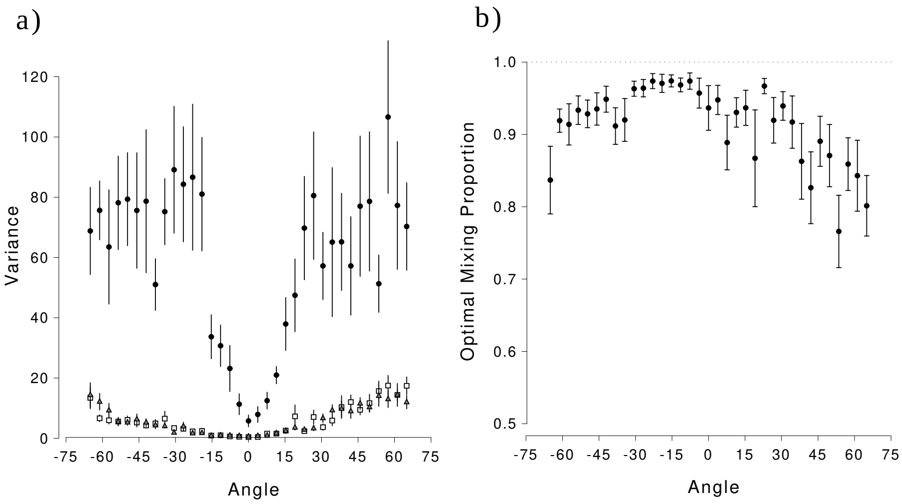

Figure 2: a) Variance of localization as a function of angle of azimuth for the three types of stimuli: visual (white squares), auditory (filled circles), visuo-auditory (filled triangles). b) Optimal mixing weights as a function of azimuth for vision, as predicted by minimum variance integration. Note that vision dominates the most straight-ahead.

To investigate the pattern of adaptation arising from a discrepancy between the visual and auditory senses, we imposed a constant spatial shift of 15° between the visual and auditory stimuli during VA trials. Only one third of the trials were VA; the V and A trials throughout the experiment could therefore be used to assess adaptation. Whereas pointing in visual-alone trials did not shift significantly as a result of the perturbation, pointing in auditory-alone trials shifted by about 40% in the direction of the displacement (Figure 3). This suggests that, as predicted by both the minimum variance and competitive integration models, the more reliable sense (vision) acts as the teaching signal for the less reliable one (audition). The minimum variance integration model predicts that vision should also adapt in the direction of audition. However, the amount of this predicted adaptation—about 10%

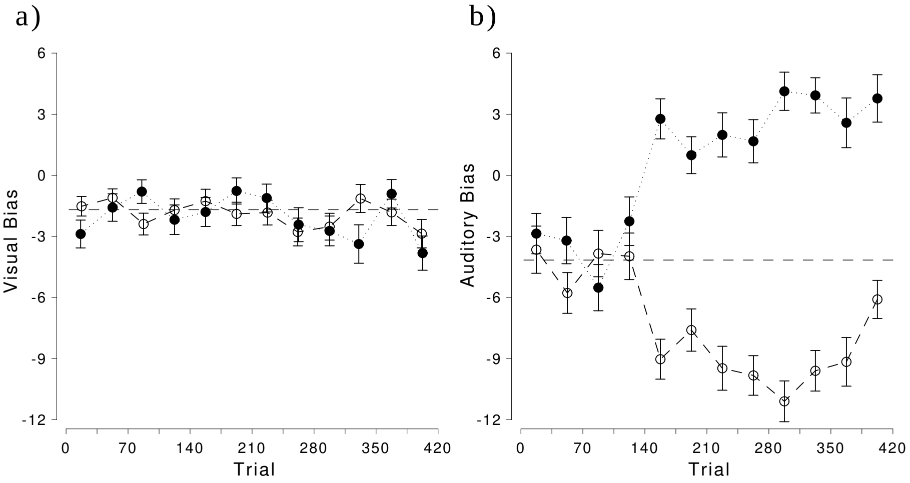

Figure 3: Adaptation as a function of trial number for a) visual and b) auditory localization. The perturbation was absent for the first 105 trials, was introduced gradually, increasing linearly, during trials 105-140, and was present in full for the remainder of the session. A baseline localization bias was computed from trials 1-105 and plotted as a dashed line. The mean $\pm$ 1 standard error bias is plotted for the group in which audition was shifted to the left (solid circles) and right (hollow circles) of vision.

---

the auditory adaptation—is within the margin of error of the experimental paradigm.

Taking the view that integration and adaptation are both related to the reliability of the sensory inputs, the finding that visual and auditory localization variance changes considerably as a function of angle of azimuth (Figure 2a) suggests that the amount of auditory adaptation may also vary as a function of azimuth. In fact, the three models of adaptation we have presented make quite distinct predictions regarding the spatial pattern of auditory adaptation:

- The *delta rule* (14) predicts that the amount of adaptation will simply be proportional to the magnitude of the displacement introduced and the duration (number of trials) of the exposure to this displacement. As both the magnitude and duration of exposure are constant throughout the azimuth, the delta rule predicts that adaptation will also be *constant* as a function of azimuth.

- The *weighted delta rule* (11) predicts that, along with magnitude and duration, the amount of auditory adaptation will also be proportional to the weighting of the visual modality. Since the data suggests that vision is weighted most heavily straight-ahead (Figure 2b), under this hypothesis auditory adaptation will be *greatest* straight-ahead.
- The *variance-weighted delta rule* (12) predicts that the amount of auditory adaptation will be proportional to the variance of auditory localization. Therefore, given the data (Figure 2a), auditory adaptation will be *least* straight-ahead.

The experimentally-obtained spatial pattern of auditory adaptation shows a pronounced reduction straight-ahead (Figure 4). These results support the variance-weighted delta rule, in which each modality adapts in proportion to its variance, in favor of the other two learning rules. Some asymmetry in adaptation is also observed, which is perhaps related to asymmetries resulting from pointing responses being made using only the right hand (Ghahramani 1995).

While these results are suggestive, further experiments are needed to further elucidate the processes of visuo-auditory integration and adaptation. So far, our results strongly argue against models in which senses are integrated stochastically. The pattern of adaptation is consistent with the variance-weighted delta rule, which in turn can be derived from minimum variance integration. These data suggest an important role for a signal coding for reliability of an input, both as a weight for multisensory integration, and as a modulator for intersensory adaptation. In the next section we examine the predictions of a Kalman filter, the dynamical extension of minimum variance integration, in the context of sensorimotor integration during arm movements.

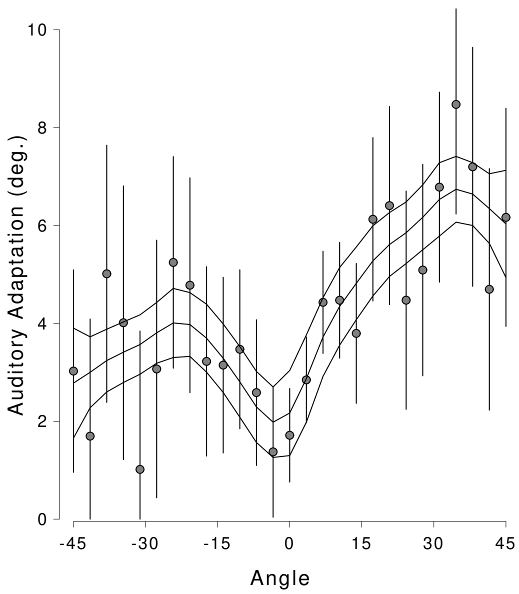

Figure 4: Auditory adaptation as a function of angle of azimuth. Means and standard errors are plotted, along with a smoothing spline fit with standard error curves.

4 Sensorimotor Integration and Internal Models

When we move our arm in the absence of visual feedback, there are three basic methods whereby the motor control system can obtain an estimate of the current state (e.g. position and velocity) of the hand. The system can make use of sensory inflow (reafference), it can make use of integrated motor outflow (dead reckoning), or it can combine these two sources of information. In order to combine sensory and motor information sources, the two problems we outlined in section 2—coordinate transformation and integration—have to be solved by the central nervous system. In section 2.2.1 we presented a simple model—the Kalman filter—which addresses both these problems in the context of linear dynamical control systems. We first outline how these problems are addressed in the Kalman filter model, before reviewing some recent results testing this model's predictions regarding the temporal propagation of errors in localizing the hand during a movement (Wolpert, Ghahramani & Jordan 1995b).

For the sensorimotor system, one key aspect of the coordinate transformation problem is that, whereas sensory signals may directly cue the location of the hand, motor outflow ("efference copy") generally does not. Knowing the sequence of torques applied to an arm, for example, does not determine its final configuration; in order to convert motor outflow into an estimate of the state of the arm, the system must make use of an *internal model* of the arm's dynamics. Specifically, there are two varieties of internal models—"forward models," which mimic the causal flow of a process by predicting its next state given the current state and the motor command, and "inverse models," which are anticausal, estimating the motor command that causes a particular state transition (Jordan 1995). The Kalman filter makes use of a forward model in order to predict the state of the arm. This motor prediction is then combined with sensory inputs according to the minimum variance integration principle (Goodwin & Sin 1984).

To examine the possibility that an internal model is indeed used in sensorimotor integration, we carried out an experiment in which subjects made arm movements in the dark (Wolpert et al. 1995b). Three experimental conditions were studied, involving the use of null, assistive and resistive force fields. Subjects gripped a planar two degree-of-freedom torque-motor-driven manipulandum (Faye 1986), while viewing virtual visual feedback projected into the plane of movement. The manipulandum was used to accurately measure the position of the subject's thumb and also, using the torque motors, to apply forces to the hand. The hand was constrained to move along a straight line passing transversely in front of the subject. Each trial started with the subject visually placing his thumb at a target square projected randomly on the movement line. The arm was then illuminated for two seconds, thereby allowing the subject to visually perceive his initial arm configuration. The light was then extinguished and the subject moved his hand left or right, as indicated by an arrow, in the absence of visual feedback. The subjects' internal estimate of hand location was assessed by asking them to localize visually the position of their hand at the end of the movement. The discrepancy between the actual and visual estimate of thumb location was recorded as a measure of the state estimation error.

The bias of the estimated location of the hand, plotted as a function of movement duration showed a consistent overestimation of the distance moved (Figure 5). This bias demonstrated two distinct phases as a function of movement duration, an initial increase reaching a peak of 0.9 cm after one second followed by a sharp transition to a region of gradual decline. The variance of the estimate also showed an initial increase during the first second of movement after which it plateaus at about 2 cm^2. External forces had distinct effects on the bias and variance propagation. Whereas the bias was increased by the assistive force and decreased by the resistive force, the variance was unaffected.

These experimental results were fully accounted for using a Kalman filter model which integrates the efferent outflow and the reafferent sensory inflow. The system dynamics of the hand was approximated by a damped (coefficient β) point mass, m, moving in one dimension acted on by a force u, combining both internal motor commands and

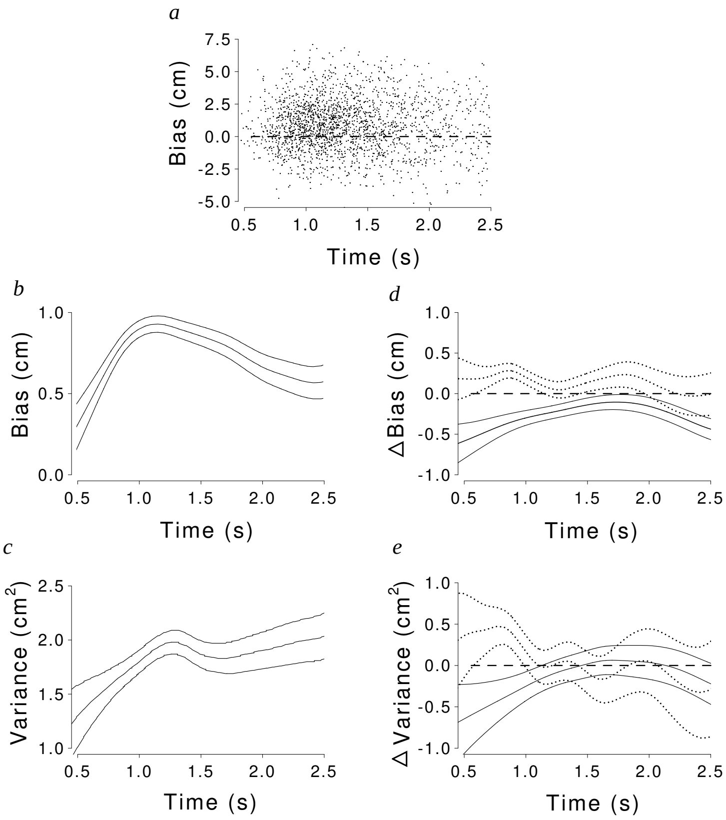

Figure 5: The raw localization bias against movement duration is shown in a) for all 8 subjects (300 trials each). There are few data points for short movement durations due to the reaction time of stopping in response to the tone—all graphs are therefore plotted from 0.5 s. b–e) show the main effect fits of a generalized additive model to the data (Hastie & Tibshirani 1990). The propagation of the (b) bias and (c) variance of the state estimate is shown, with outer standard error lines, against movement duration. The differential effects on (d) bias and (e) variance of the external force, assistive (dotted lines) and resistive (solid lines), are also shown relative to zero (dashed line). A positive bias represents an overestimation of the distance moved. The difference in variance propagation between the resistive and assistive fields was not significant over the movement; the difference in bias was significant at the $p = 0.05$ level. Reprinted with permission from Wolpert, Ghahramani and Jordan (1995b).

external forces. Representing the state of the hand at time t as $\mathbf{x}(t)$ (a 2×1 vector of position and velocity), the system dynamic equations can be written in the general form of $\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + Bu(t) + \mathbf{w}(t)$ where $A = \begin{bmatrix} 0 & 1 \\ 0 & -\beta/m \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ 1/m \end{bmatrix}$ and the vector $\mathbf{w}(t)$ represents the process of white noise. The system has an observable output, $\mathbf{y}(t)$, representing the proprioceptive signals (e.g. from muscle spindles and joint receptors), which is linked to the actual hidden state $\mathbf{x}(t)$ by $\mathbf{y}(t) = C\mathbf{x}(t) + \mathbf{v}(t)$ where the vector $\mathbf{v}(t)$ represents the output white noise. We assume that this system is fully observable and choose C to be the identity matrix. At time $t = 0$ the subject was given full view of his arm and, therefore, started with an estimate $\hat{\mathbf{x}}(0) = \mathbf{x}(0)$ with zero bias and variance—i.e. vision calibrated the system. At this time the light was extinguished and the subject had to rely on the inputs and outputs to estimate the system's state. The Kalman filter, using a model of the system $\hat{A}$, $\hat{B}$ and $\hat{C}$, provides an optimal linear estimator of the state given by

$$\dot{\hat{\mathbf{x}}}(t) = \underbrace{\hat{A}\hat{\mathbf{x}}(t) + \hat{B}u(t)}_{\text{Forward model}} + \underbrace{K(t)[\mathbf{y}(t) - \hat{C}\hat{\mathbf{x}}(t)]}_{\text{Sensory correction}}$$

where $K(t)$ is the recursively updated Kalman gain matrix (Figure 6a). This state estimate combines an estimate from the internal model of the system dynamics together with a sensory correction. The relative contributions of the internal simulation and sensory correction processes to the final estimate are modulated by the Kalman gain matrix so as to provide minimum variance state estimates. We use this state update equation to model the bias and variance propagation and the effects of the external force. The parameters in the simulation, β, m and u were chosen based on the mass of the arm and the observed relationship between time and distance traveled.

By making particular choices for the parameters of the Kalman filter, we are able to simulate dead reckoning, sensory inflow-based estimation, and forward model-based sensorimotor integration. Moreover, to accommodate the observation that subjects generally tend to overestimate the distance that their arm has moved, we set the gain that couples force to state estimates to a value that is larger than its veridical value. This setting is consistent with independent data that subjects tend to under-reach in pointing tasks, suggesting an overestimation of distance traveled (Soechting & Flanders 1989). All other components of the internal model were set to their veridical values.

Simulations of the Kalman filter demonstrated the two distinct phases of bias propagation observed (Figure 6). By overestimating the force acting on the arm the forward model overestimates the distance traveled, an integrative process eventually balanced by the sensory correction. The model also captured the differential effects on bias of the externally imposed forces. By overestimating an increased force under the assistive condition, the bias in the forward model accrues more rapidly and is balanced by the sensory feedback at a higher level. The converse applies to the resistive force. The pattern of variance propagation was also captured by the model. During the early part of the movement, because of the initial visual calibration the current state estimate resulting from the forward model is accurate, and therefore the sensorimotor integration process weights it more heavily. However, in the later stages of the movement, when the current state estimate is less accurate, the sensory feedback must be relied upon to correct for inaccuracies in the forward model. In the Kalman filter, the relative weighting shifts from the forward model towards sensory feedback over the first second of movement and then remains approximately constant resulting in the asymptote of the variance propagation. In accord with the experimental results the model predicts no change in variance under the two force conditions.

These results show that the Kalman filter is able to reproduce the propagation of the bias and variance of estimated position of the hand as a function of both movement duration and external forces. The model also simulates the interesting and novel empirical result that while the variance asymptotes, the bias peaks after about one second and then gradually declines. This behavior is a consequence of a trade off between the inaccuracies accumulating in the internal simulation of the arm's dynamics and the feedback of actual sensory information. Simple models which do not trade off the contributions of a forward

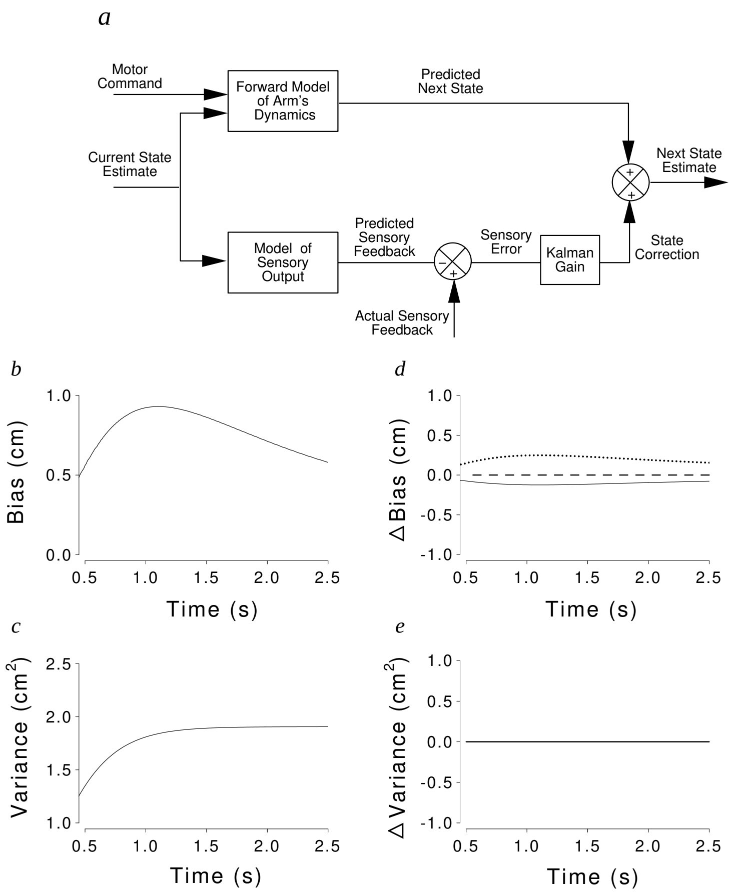

Figure 6: a) The Kalman filter model comprises two processes. The first (upper part) simulates the arm's dynamics using the motor command and the current state estimate to predict the next state estimate. The second process (lower part) uses the difference between expected and actual sensory feedback to correct the forward model state estimate. The relative weighting of these two processes is mediated through the Kalman gain. b–e) Simulated bias and variance propagation from the Kalman filter model of the sensorimotor integration process, in the same representation and scale as Figure 5 (b–e). Reprinted with permission from Wolpert, Ghahramani and Jordan (1995b).

model with sensory feedback, such as those based purely on sensory inflow or on motor outflow, are unable to reproduce the observed pattern of bias and variance propagation. The ability of the Kalman filter to parsimoniously model this data suggests that the processes embodied in the filter, namely internal simulation through a forward model together with sensory correction, are likely to be embodied in the sensorimotor integration process.

5 Relevance to Neurophysiology

One candidate for the neural system subserving both the integration of visual and auditory inputs and the production of orienting movements to such inputs is the superior colliculus (SC). The superior colliculus, and its non-mammalian homologue, the optic tectum, is a layered midbrain structure in which the superficial layers receive visual inputs both directly from the retina and from visual cortex, and the deep layers receive visual, somatosensory, auditory and motor-related inputs (Wickelgren 1971, Harris et al. 1980, Stein & Meredith 1993). Over 50% of neurons in the deep layer are multi-sensory, with visuo-auditory being the most common combination (30% of total; Stein & Meredith 1993). It is important to note that multisensory convergence seems to take place at the deep layer neuron itself, most of whose inputs are unimodal (Wickelgren & Sterling 1969). The outputs of the superior colliculus project to brain stem and spinal cord areas directly involved in positioning the peripheral sense organs, playing an important role in orienting the eyes, head, limbs and, in species that can move them, ears and whiskers (Harris et al. 1980, Sparks & Nelson 1987, DuLac & Knudsen 1990, Guitton & Munoz 1991, Stein & Meredith 1993).

Knudsen and colleagues have extensively studied adaptation to visuomotor and visuo-auditory displacements and their effects on the neural representations of space in the optic tectum of the barn owl. Their results have shown that prismatically perturbing visual inputs, while barely modifying visual localization, induced significant adaptation of auditory localization (Knudsen & Knudsen 1989a, Knudsen & Knudsen

1989b). Furthermore, blind-reared owls developed highly abnormal maps of auditory space in the optic tectum (Knudsen et al. 1991). Our findings are consistent with these results, again suggesting that the registration of visual and auditory maps is largely determined by visual experience. Recently, it has been found that adaptation of the auditory map in the optic tectum can be attributed to changes in one of its inputs, the inferior colliculus (Brainard & Knudsen 1993). Further research needs to be done to determine the signal driving adaptation in the inferior colliculus (cf. the model proposed by Pouget, Deffayet & Sejnowski 1995).

Our computational models and experimental results suggest that sensory inputs in an area such as SC may be weighted by a measure of their reliability. The reliability of a sensory input must therefore somehow be coded neurally, along with the input itself. One possibility is that the firing rate of a neuron in the spatial map could be proportional to that neuron's "confidence" that there is a stimulus in its receptive field. Under this hypothesis, there are two explanations for the finding that animals often orient to a locus in between visual and auditory stimuli presented simultaneously at different locations (Stein et al. 1989). First, the two distinct loci of activity may merge into one intermediate locus within the collicular map. Second, the actual integration of signals may occur at a later motor stage, whose units have large receptive fields in the collicular map. An alternative to this explicit rate-coding hypothesis for reliability of sensory inputs is that reliability is coded implicitly in the neural architecture. For example, the size of receptive fields could be related both to the variance in localization and to the rate of plasticity. (Note also that receptive fields are larger in the periphery, where we found greater adaptation.) More detailed neurophysiologically-based models of the colliculus may provide links between the computational, psychological and neural levels of understanding the problem of visuo-auditory integration.

Finally, the state estimation paradigm we used in the study of sensorimotor integration during arm movements provides a framework to study integration process in both normal and patient populations. For example, the specific predictions of the sensorimotor integration model can be tested in both patients with sensory neuropathies, who lack proprioceptive reafference, and in patients with damage to the cerebellum, a proposed site for a forward model (Miall et al. 1993). Here again, the computational model will hopefully provide a reference point for interpreting new behavioral and neurological results.

References

Abidi, M. & Gonzalez, R. (1992). *Data Fusion In Robotics and Machine Intelligence*, Academic Press, San Diego, CA.

Attneave, F. (1954). Some informational aspects of visual perception, *Psych. Review* **61**: 183–193.

Barlow, H. (1961). Three points about lateral inhibition, *in* W. Rosenblith (ed.), *Sensory Communication*, MIT Press, Cambridge, MA, pp. 782–786.

Becker, S. & Hinton, G. (1992). A self-organizing neural network that discovers surfaces in random-dot stereograms, *Nature* **355**: 161–163.

Bennett, B. M., Hoffman, D. D. & Prakash, C. (1989). *Observer Mechanics*, Academic Press, San Diego.

Brainard, M. & Knudsen, E. (1993). Experience-dependent plasticity in the inferior colliculus: A site for the visual calibration of the neural representation of auditory space in the barn owl, *J. Neuroscience* **13**(11): 4589–4608.

Cover, T. & Thomas, J. (1991). *Elements of Information Theory*, Wiley, New York.

DuLac, S. & Knudsen, E. (1990). Neural maps of head movement vector and speed in the optic tectum of the barn owl, *J. Neurophysiology* **63**(1): 131–146.

Durrant-Whyte, H. F. (1988.). *Integration, coordination, and control of multi-sensor robot systems*, Kluwer Academic Publishers, Boston.

Faye, I. (1986). *An impedence controlled manipulandum for human movement studies*, MS Thesis, MIT Dept. Mechanical Engineering, Cambridge, MA.

Flash, T. & Hogan, N. (1985). The co-ordination of arm movements: An experimentally confirmed mathematical model, *J. Neuroscience* **5**: 1688–1703.

Ghahramani, Z. (1995). *Computation and Psychophysics of Sensorimotor Integration*, Ph.D. Thesis, Dept. of Brain and Cogntive Sciences, Massachusetts Institute of Technology, Cambridge, MA.

Ghahramani, Z., Wolpert, D. M. & Jordan, M. I. (1995). Computational principles of multisensory integration: Studies of adaptation to novel visuo-auditory remappings, *Society For Neuroscience Abstracts* **21**(1-3): 1181.

Gibson, J. J. (1961). Ecological optics, *Vision Research* **1**: 253–262.

Goodwin, G. & Sin, K. (1984). *Adaptive filtering prediction and control*, Prentice-Hall.

Guitton, D. & Munoz, D. (1991). Control of orienting gaze shifts by the tectoreticulospinal system in the head-free cat. I. Identification, localization, and effects of behavior on sensory responses , *J. Neurophysiology* **66**: 1605–1623.

Harris, L., Blakemore, C. & Donaghy, M. (1980). Integration of visual and auditory space in the mammalian superior colliculus, *Nature* **288**: 56–59.

Hastie, T. & Tibshirani, R. (1990). *Generalized Additive Models*, Chapman and Hall, London.

Hertz, J., Krogh, A. & Palmer, R. (1991). *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City, CA.

Hogan, N. (1984). An organizing principle for a class of voluntary movements, *J. Neuroscience* **4**: 2745–2754.

Howard, I. P. & Templeton, W. B. (1966). *Human Spatial Orientation*, Wiley, New York.

Ito, M. (1984). *The Cerebellum and Neural Control*, Raven Press, New York.

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*, Academic Press, New York.

Jordan, M. I. & Rumelhart, D. (1992). Forward models: Supervised learning with a distal teacher, *Cognitive Science* **16**: 307–354.

Kalman, R. & Bucy, R. S. (1961). New results in linear filtering and prediction, *Journal of Basic Engineering (ASME)* **83D**: 95–108.

Kawato, M., Furawaka, K. & Suzuki, R. (1987). A hierarchical neural network model for the control and learning of voluntary movements, *Biological Cybernetics* **56**: 1–17.

Knudsen, E., Esterly, S. & DuLac, S. (1991). Stretched and upside-down maps of auditory space in the optic tectum of blind-reared owls; Acoustic basis and behavioral correlates, *J. Neuroscience* **11**(6): 1727–1747.

Knudsen, E. & Knudsen, P. (1989a). Vision calibrates sound localization in developing barn owls, *J. Neuroscience* **9**(9): 3306–3313.

Knudsen, E. & Knudsen, P. (1989b). Visuomotor adaptation to displacing prisms by adult and baby barn owls, *J. Neuroscience* **9**(9): 3297–3305.

Linsker, R. (1986). From basic network principles to neural architecture: Emergence of orientation selective cells, *Proceedings of the National Academy of Sciences USA* **83**: 8390–8394.

Marr, D. (1982). *Vision*, Freeman, New York.

Miall, R., Weir, D., Wolpert, D. M. & Stein, J. (1993). Is the cerebellum a Smith Predictor?, *Journal of Motor Behavior* **25(3)**: 203–216.

Middlebrooks, J. C. & Green, D. M. (1991). Sound localization by human listeners, *Ann. Rev. Psychol.* **42**: 135–159.

Mills, A. W. (1958). On the minimum audible angle, *Journal of the Acoustical Society of America* **30**: 237–246.

Nakayama, K. & Shimojo, S. (1992). Experiencing and perceiving visual surfaces, *Science* **257**: 1357–1363.

Nelson, W. (1983). Physical principles for economies of skilled movements, *Biological Cybernetics* **46**: 135–147.

Pouget, A., Deffayet, C. & Sejnowski, T. (1995). Reinforcement learning predicts the site of plasticity for auditory r remapping in the barn owl, *in* G. Tesauro, D. Touretzky & T. Leen (eds), *Advances in Neural Information Processing Systems 7*, MIT Press, Cambridge, MA.

Richards, W. (1988). *Natural Computation*, MIT Press, Cambridge, MA.

Rumelhart, D. & McClelland, J. (1986). *Parallel distributed processing*, MIT press, Cambridge, Mass.

Shannon, C. (1948). A mathematical theory of communication, *Bell Systems Technical Journal* **27**: 379–423,623–656.

Soechting, J. & Flanders, M. (1989). Sensorimotor representations for pointing to targets in three-dimensional space, *J. Neurophysiology* **62**: 582–594.

Sparks, D. & Nelson, J. (1987). Sensory and motor maps in the mammalian superior colliculus, *Trends in Neuroscience* **10**(8): 312–317.

Stein, B. & Meredith, M. (1993). *The Merging of the Senses*, MIT Press, Cambridge, MA.

Stein, B., Meredith, M., Honeycutt, W. & McDade, L. (1989). Behavioral indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli, *Journal of Cognitive Neuroscience* **1**: 12–24.

Uno, Y., Kawato, M. & Suzuki, R. (1989). Formation and control of optimal trajectories in human multijoint arm movements: Minimum torque-change model, *Biological Cybernetics* **61**: 89–101.

Wickelgren, B. (1971). Superior colliculus: Some receptive field properties of bimodally responsive cells, *Science* **173**: 69–71.

Wickelgren, B. & Sterling, P. (1969). Influence of visual cortex on receptive fields in the superior colliculus of the cat, *J. Neurophysiology* **32**: 16–23.

Widrow, B. & Hoff, M. (1960). Adaptive switching circuits, *IRE WESCON Convention Record*, Vol. 4, IRE, New York, pp. 96–104.

Wolpert, D. M., Ghahramani, Z. & Jordan, M. I. (1995a). Are arm trajectories planned in kinematic or dynamic coordinates? An adaptation study , *Experimental Brain Research* **103**(3): 460–470.

Wolpert, D. M., Ghahramani, Z. & Jordan, M. I. (1995b). An internal model for sensorimotor integration, *Science* **269**: 1880–1882.