Computation and Psychophysics of Sensorimotor Integration

by

Zoubin Ghahramani

B.S.E., Computer Science, University of Pennsylvania (1990)
B.A., Cognitive Science, University of Pennsylvania (1990)

Submitted to the Department of Brain and Cognitive Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1995

©Massachusetts Institute of Technology 1995.
All rights reserved.

Computation and Psychophysics of Sensorimotor Integration

by

Zoubin Ghahramani

Submitted to the Department of Brain and Cognitive Sciences
on July 28, 1995, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

All higher organisms are able to integrate information from multiple sensory modalities and use this information to select and guide movements. In order to do this, the central nervous system (CNS) must solve two problems: (1) Converting information from distinct sensory representations into a common coordinate system, and (2) integrating this information in a sensible way. This dissertation proposes a computational framework, based on statistics and information theory, to study these two problems. The framework suggests explicit models for both the coordinate transformation and integration problems, which are tested through human psychophysics.

The experiments in Chapter 2 suggest that: (1) Spatial information from the visual and auditory systems is integrated so as to minimize the variance in localization. (2) When the relation between visual and auditory space is artificially remapped, the spatial pattern of auditory adaptation can be predicted from its localization variance. These studies suggest that multisensory integration and intersensory adaptation are closely related through the principle of minimizing localization variance. This principle is used to model sensorimotor integration of proprioceptive and motor signals during arm movements (Chapter 3). The temporal propagation of errors in estimating the hand's state is captured by the model, providing support for the existence of an internal model in the CNS that simulates the dynamic behavior of the arm.

The coordinate transformation problem is examined in the visuomotor system, which mediates reaching to visually-perceived objects (Chapter 4). The pattern of changes induced by a local remapping of this transformation suggests a representation based on units with large functional receptive fields. Finally, the problem of converting information from disparate sensory representations into a common coordinate system is addressed computationally (Chapter 5). An unsupervised learning algorithm is proposed based on the principle of maximizing mutual information between two topographic maps. What results is an algorithm which develops multiple, mutually-aligned topographic maps based purely on correlations between the inputs to the different sensory modalities.

Thesis Supervisor: Michael I. Jordan
Title: Professor

Thesis Co-Supervisor: Tomaso Poggio
Title: Uncas and Helen Whitaker Professor

Acknowledgments

I thank Daniel Wolpert for reasons too numerous to mention here. Without his mentorship, encouragement, patience and creativity there would be no "psychophysics" in this thesis. He made hard work enjoyable and exciting, and I doubt I will find such a combination of friend and collaborator in the future.

I thank Michael Jordan for providing an unparalleled training environment. I learned more in his grueling three-hour lab meetings than in the rest of my graduate coursework combined. Without his mentorship there would be no "computation" in this thesis.

I thank Dick Held, Emilio Bizzi and Tomaso Poggio, for kindly serving as members of my thesis committee. Peter Dayan provided helpful comments on the manuscript and I benefitted from engaging discussions with Geoff Hinton, who also made it possible for me to write this thesis without worrying about searching for a postdoctoral position.

All the members of Jordan lab provided an excellent environment for research. I especially thank Flip Sabes for critical comments on the manuscript and friendly moral support during the last weeks of writing, Lawrence Saul for tutorials on statistical mechanics, and Tommi Jaakkola for letting me measure his head. Carlotta Domeniconi provided many hours of assistance in conducting experiments. Adee Matan, who earned Jordan lab membership by always using our computers, deserves a special thanks for keeping track of my progress and mental sanity.

I thank David Poeppel for egging me on to defend early and for being just as stressed as I was during our last few days. John Houde was my dietary advisor during the thesis and travel companion throughout graduate school. I thank James Thomas for distracting me late at night with horror stories and good music while I was trying to write my dissertation. Thanks to Gregg Solomon for advice on writing Acknowledgment sections.

I thank Jan Ellertsen for guiding me through the tortuous road of academic requirements. I am grateful to Marney Smyth for dedicating many hours to helping me with slides and figures. Ellie Bonsaint provided superb administrative support throughout my graduate education, and Pat Claffey tracked down many obscure articles for me.

I have enjoyed graduate school immensely, mostly due to the wonderful environment provided by the students in the program. I thank each of them for their individual gift in making this department unique. I am grateful to the McDonnell-Pew Foundation for supporting my studies in this department.

I especially want to thank Azita Ghahramani for letting me crash at her place for 5 years while doing a PhD, and for being the most wonderful sister and housemate, and Monica Biagioli for moral support and making my last few months here very special.

Biographical note

Although my family is originally from Shiraz, Iran, I was born in Moscow on February 8^{th}, 1970. After four years in Russia, my family moved back to Iran for one year, and then to Madrid, Spain. I lived in Spain and attended the American School of Madrid from 1976 until my high school graduation in 1986. I then went to Philadelphia to study at the University of Pennsylvania, where I obtained a B.A. in Cognitive Science and a B.S.E. in Computer Science. In 1990 I entered the doctoral program in Brain and Cognitive Sciences at MIT.

To my father,

for all the joy he brought me.

Contents

1 Introduction **19**
1.1 Outline of the Thesis 21

I Integration **23**

2 Integration and Adaptation of Visual and Auditory Maps **25**
2.1 Introduction 25
2.2 Background 27
2.2.1 Psychophysics 27
2.2.2 Neuroscience 32
2.3 The Computational Model 35
2.3.1 Integration 35
2.3.2 Adaptation 39
2.3.3 Related Models 41
2.3.4 Summary 43
2.4 Overview of the Experiments 43
2.5 Experiment 1: Localization of Visual, Auditory, and Visuo-auditory Stimuli 44
2.5.1 Method 45
2.5.2 Results 48
2.5.3 Discussion 49
2.6 Experiment 2: Adaptation to a Visuo-Auditory Remapping 54
2.6.1 Method 54
2.6.2 Results 55
2.6.3 Discussion 56
2.7 Experiment 3: Adaptation to Visuo-Auditory Variance 62
2.7.1 Method 62
2.7.2 Results 63
2.7.3 Discussion 64
2.8 Experiment 4: Generalization of the Visuo-Auditory Map 67
2.8.1 Method 68

2.8.2 Results . . . 68
2.8.3 Discussion . . . 71
2.9 Controls . . . 72
2.9.1 Alternative Cues to Auditory Stimulus Location . . . 72
2.9.2 Pointing with the Left Hand . . . 73
2.10 Discussion . . . 74
2.10.1 Empirical findings . . . 74
2.10.2 Implications . . . 77
2.10.3 Directions for future work . . . 78
2.10.4 Conclusion . . . 78

3 An Internal Model for Sensorimotor Integration 79
3.1 Introduction . . . 79
3.2 Experiment: Propagation of Errors in Sensorimotor Integration . . . 80
3.3 Appendix A: Paradigm . . . 84
3.4 Appendix B: Simulation . . . 85

II Coordinate Transformations 87

4 Representation of the Visuomotor Coordinate Transformation 89
4.1 Introduction . . . 89
4.2 The Visuomotor Coordinate Transformation . . . 90
4.2.1 Spatial Generalization . . . 90
4.2.2 Contextual Generalization . . . 92
4.2.3 Experimental Aims and Overview . . . 93
4.3 Experiment 1: Visuomotor Generalization to a One-Point Displacement 94
4.3.1 Method . . . 95
4.3.2 Results . . . 98
4.3.3 Discussion . . . 100
4.4 Experiment 2: Visuomotor Generalization to a Two-Point Displacement 101
4.4.1 Method . . . 102
4.4.2 Results . . . 103
4.4.3 Discussion . . . 104
4.5 Experiment 3: Contextual Generalization of the Visuomotor Map A . 106
4.5.1 Method . . . 107
4.5.2 Results . . . 109
4.6 Experiment 4: Contextual Generalization of the Visuomotor Map B . 111
4.6.1 Method . . . 111
4.6.2 Results . . . 112
4.6.3 Discussion . . . 113
4.7 General Discussion . . . 114
4.8 Conclusion . . . 119

5 Learning Coordinate Transformations through Mutual Information 131
5.1 Introduction 131
5.2 Information-theoretic Unsupervised Learning 132
5.2.1 Unsupervised Learning 132
5.2.2 Information Theory 134
5.2.3 Previous approaches 135
5.3 Topographic Mutual Information 137
5.3.1 The cost function 137
5.3.2 The learning algorithm 141
5.4 Experiment: Polar and Cartesian Coordinates 142
5.5 Discussion 146

6 Conclusion 155

List of Figures

2-1 Visuo-auditory experimental setup 46
2-2 Experimental paradigm 48
2-3 Average localization bias plotted as a function of stimulus location . . 50
2-4 Average variance of localization plotted as a function of stimulus location 51
2-5 Optimal mixing proportion for vision predicted by minimum variance integration 53
2-6 Adaptation as a function of trial number 56
2-7 Auditory adaptation as a function of target location 57
2-8 Visual adaptation as a function of target location 58
2-9 Auditory adaptation as a function of location 61
2-10 Variance of localization as a function of session 64
2-11 Variance of localization as a function of stimulus location 65
2-12 Pattern of generalization 69
2-13 Time course of generalization 70
2-14 Bias as a function of stimulus location for left-hand control 75
2-15 Variance as a function of stimulus location for left-hand control . . . 76

3-1 The effect of movement duration and external forces on the propagation of bias and variance of the state estimate 82
3-2 Simulated bias and variance propagation from a Kalman filter model of the sensorimotor integration process 83
3-3 Experimental apparatus combining virtual visual feedback and planar manipulandum 85

4-1 Schematic of some possible patterns of generalization 120
4-2 Visuo-motor perturbation setup 121
4-3 Experimental paradigm 122
4-4 Targets and pre-exposure pointing locations for Experiment 1 123
4-5 Target acquisition time as a function of trial during the exposure phase 123
4-6 Average change in pointing for the control group 124
4-7 Average change in pointing for the x-shift group 124
4-8 Average change in pointing for the y-shift group 125

4-9 Schematic of the perturbation and the layout of the 11 targets for Experiment 2 125
4-10 Targets and pre-exposure pointing locations for Experiment 2 126
4-11 Target acquisition time as a function of trial during the exposure phase 126
4-12 Average change in pointing for the Experiment 2 control group 126
4-13 Average change in pointing for the y-shift group 127
4-14 Schematic of the exposure phase of Experiment 3 128
4-15 Changes in pointing as a function of starting position in Experiment 3 129
4-16 Adaptation in the x direction plotted as a function of starting point . 129
4-17 Change in pointing as a function of starting position in Experiment 4 130

5-1 The basic model for multisensory integration 138
5-2 Learning curves for the non-topographic and topographic mutual information algorithms 143
5-3 Learning with the non-topographic mutual information algorithm . . 144
5-4 Learning with the topographic mutual information algorithm 145

List of Tables

2.1 Summary of model predictions. 60
2.2 Summary of alternative cues control experiment. 73

4.1 Summary of the two-factor within-subject ANOVAs for Experiment 1 99
4.2 Summary of the two-factor within-subject ANOVAs for Experiment 2 104
4.3 Summary of the two-factor within-subject ANOVAs for Experiment 3 110
4.4 Summary of the two-factor within-subject ANOVAs for Experiment 4 112

Chapter 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 onset, this problem seems formidable. Information arriving into each sense codes for disparately different aspects of the environment: Audition senses changes in pressure on the eardrum, vision detects photons on the retina, the sense of smell recognizes individual molecules in the olfactory bulb. The central nervous system accomplishes the astonishing feat 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. The simple act of reaching, for example, may require convergence of information from the visual, proprioceptive, and motor systems.

The principles underlying sensorimotor integration—the ability to integrate information from multiple sensory and motor systems—are the topic of this thesis. Two goals are set regarding this topic. The first goal is to build a computational theory of sensorimotor integration in the tradition of Marr (1982). While each sensory modality and motor subsystem is distinct in its functioning, there are commonalities in the problem of integrating multiple sources of information that can be captured within a computational framework. The second goal is to test, through psychophysical experiments, the simple models of sensorimotor integration that arise from this computational framework. The predictions of these models often transcend particular systems; we therefore study integration in three distinct sensorimotor systems. Chapter 2 investigates the integration of the visual and auditory systems when we localize an object; Chapter 3 focuses on the sensorimotor system involved in tracking the hand during movement; Chapter 4 focuses on the transformation from the visual location of an object to the pointing response required to reach it.

The two goals of this thesis can be formulated as answers to two fundamental questions:

What is the problem of sensorimotor integration? The central nervous system (CNS) receives information from multiple sensory modalities and integrates these sources into unified percepts and motor acts.[1] Can the problem of integrating multiple information sources be formulated abstractly? I propose that the answer is Yes—the problem can be posed, in a meaningful way, within the closely related computational frameworks of statistics and information theory.

The problem of integrating multiple sources can be decomposed into two problems: How to convert information in multiple disparate representations into a common representation appropriate for integration, and how to combine information that is already in a common representation in an optimal way. The second problem—*integration*—is the focus of the Part I of this thesis. The first and perhaps more difficult problem—*coordinate transformation*—is the focus of Part II.

How does the CNS solve the problem of sensorimotor integration? One can derive several models of integration based on different computational criteria for optimality. The models capture simple intuitive ways in which multiple information sources could be combined in the CNS. For example, less reliable sources could be ignored in favor of more reliable sources. Can these models appropriately characterize sensorimotor integration in the CNS? To answer this question affirmatively, converging lines of psychophysical evidence are required. The series of experiments in Chapters 2 and 3, are designed to test these models in two distinct sensorimotor systems.

There may be no single principle underlying the integration of information from multiple sources in the brain. We start, however, with the hypothesis that there is such a principle, and put this hypothesis to test experimentally. The disadvantage of this theory-driven approach is that we may easily be wrong. The advantage is that, in the process of testing our hypothesis, the questions being addressed are clarified and extensions of the models are suggested.

In viewing the human sensorimotor system from a computational perspective, it is natural to ask what the advantages of designing a multisensory system may be.[2] The study of robotics suggests three principal advantages of combining multiple information sources (Abidi and Gonzalez, 1992):

- Multiple sensors provide *redundancy*, which can reduce the overall uncertainty of sensory estimates and increase the reliability in the case of sensor failure.

[1] We will review the evidence that such integration does indeed occur in Chapter 2.

[2] We use *multisensory integration* to refer to combining information from different sensory modalities. We use *sensorimotor integration* to refer to using this information for the production of movement. The boundary between multisensory and sensorimotor integration is blurred by the fact that movement gives rise to both reafferent sensory signals and copies of the motor efference signal (see Chapter 3). We occasionally use the two terms interchangeably to refer to the integration of signals, regardless of their sensory or motor origin.

- *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).

- *More timely information* may be obtained through parallelism, as each sensor may have a different latency.

Although all three factors may have played a role selecting for multisensory systems in biological organisms, we focus on the first factor. In Chapters 2 and 3 we show how it is possible to quantify exactly the reduction in uncertainty arising from integrating multiple sensors.

The robotics literature also distinguishes between different levels of multisensory integration (Abidi and Gonzalez, 1992). If the inputs from the different sensors are synchronized, and in the same coordinate system, then they can be combined at the *signal level*. For example, two noisy radar signals can be averaged to yield a clearer signal. If the inputs are not necessarily in the same coordinate frame but provide distinct features at a higher level of representation, then they can be combined at the *feature level*. For example, using two hands a robot could feel different parts of a object—allowing it to recognize the whole object. Finally, if the information from each sensor is represented as a logical proposition or probability, then the senses can be combined at the *symbol level*. For example, a distributed system consisting of multiple robots could attempt to make a decision by pooling the opinions of each robot into a single opinion. The integration processes we study in this thesis can be considered to be at the two lower, signal and feature, levels of this hierarchy. The focus is not on the cognitive components of combining information from different sources but on low level perceptuomotor processes. Thus, although subjects were consciously unaware of any discrepancies between inputs from different sensory modalities, their perceptuomotor system reflected the effects of such discrepancies.

1.1 Outline of the Thesis

Part I of the thesis (Chapters 2 & 3) focuses on the problem of integrating information from different sensory modalities. A computational framework for multisensory integration is derived from statistical estimation theory in Chapter 2. In this framework, the problem of integrating multiple modalities is closely tied to the problem of adapting to discrepancies between modalities. A series of four experiments is described in which the integration and adaptation of visual and auditory maps is examined using a paradigm in which subjects point to visual, auditory, or combined visuo-auditory stimuli. The results of these experiments are compared to the predictions of the different computational models of intersensory integration and adaptation.

The statistical estimation framework presented in Chapter 2 can be used to derive a model for the integration of proprioceptive and motor efference signals during arm

movement. In this model, the CNS obtains an estimate of the position of the hand by combining the outputs of an internal model, which simulates the dynamics of the arm using the motor efference, and a sensory correction based on proprioception. Chapter 3 tests the predictions of this model in a sensorimotor integration task in which subjects estimate the location of the hand after varying distance movements under external forces. This paradigm is used to test for the existence of an internal model in the CNS.

In the second part of this thesis I focus on the problem of coordinate transformations. Chapter 4 examines how the visuomotor coordinate transformation, which converts the visual locations of objects into coordinates appropriate for movement, is represented. One way in which this question can be addressed is by examining the patterns of generalization that emerge from a limited remapping. Two questions concerning visuomotor generalization are examined: (1) What changes in pointing behavior emerge over the workspace subsequent to a local remapping? (2) Can the visuomotor system be taught to map one location in visual space to two different finger positions depending on the starting point of movement—and if so, how does this remapping generalize to other starting points?

Chapter 5 addresses the problem of learning coordinate transformations. Specifically, how does the CNS extract the information that is common to several sensory inputs, each coded in its own representational system, and convert this information into a common representational system ? To address this problem at the computational level, an unsupervised learning algorithm is proposed. This algorithm is derived from an information-theoretic principle which states that a common representation can be obtained by maximizing the mutual information between the sensory modalities (Becker and Hinton, 1992), while maintaining a topographic relation between the modalities. What results is an algorithm which learns multiple, mutually-aligned topographic maps based on correlations between the inputs to the different sensory modalities.

Part I

Integration

Chapter 2

Integration and Adaptation of Visual and Auditory Maps

2.1 Introduction

A problem that all higher organisms face is how to integrate information from multiple sensory modalities. Information of central behavioral relevance, such as the location of a predator, the body's orientation, or the linguistic identity of an utterance, often arrives from different sensory modalities. When the senses give conflicting information, whether as a result of inherent distortions, noise, or experimental manipulation, the central nervous system (CNS) is faced with the problem of resolving this disagreement. How the CNS integrates sensory information is the primary question posed in this chapter. When the disagreement between two senses persists over time it is usually a cue that one of the senses is miscalibrated; the CNS often resolves this long-term discrepancy through a process of recalibration. This recalibration process, also known as adaptation or remapping, is the second topic of this chapter.

The basic thesis is that (1) the processes of intersensory integration and intersensory adaptation are inextricably linked, and that (2) there is an underlying and sensible principle that can characterize both processes. The principle states that information is integrated in proportion to some measure of the reliability of each source. A more reliable source, such as vision for locating an object straight-ahead or audition for perceiving the utterance of a speaker, is weighted more heavily than other sources. When there is a long-term discrepancy, the reliability of each source is used to determine how much it should adapt. The less reliable source is adapted proportionately more than the more reliable source, eventually reaching agreement at some middle ground.

The principle of weighting more reliable sources can be derived formally from the statistical theory of estimation. In estimation theory, the goal is to estimate a set of unknown parameters from noisy measurements of some observable variables and a statistical model relating the parameters to the observables. For example, the parameter may be the chemical composition of a star and the observables spectral measurements, or the parameter may be the location of an underwater fault and the observables sonar readings. The principle unifying these diverse estimation problems is that, given a model relating the parameters and the variables and some estimate of the noise in each process, there is an optimal way to fuse multiple sources of information. The optimal fusion combines all the sources, each weighted by its reliability, defined as the inverse of the variance of the noise in that source. The practical applications of estimation theory are as widespread as is suggested by its generality.

In this chapter we examine integration and adaptation of visual and auditory information in humans from the computational framework of estimation theory. Both vision and audition provide information on the locations of objects in the environment. It is known that each modality maintains separate maps of space (Konishi et al., 1988).[1] However, it is also clear that in certain areas of the CNS, information is integrated from both modalities into a common map (Wickelgren, 1971). These multisensory areas play a central role in basic motor responses, such as saccadic eye movements or orienting head movements (Sparks and Nelson, 1987). The problem of multisensory integration addressed in this chapter is therefore closely tied to the problem of selecting a single motor response to multiple sensory stimuli.

The outline of the chapter is as follows. In the next section I provide selective reviews of the psychophysics and neuroscience literatures on multisensory integration and visual and auditory adaptation. In section 2.3 I present a computational model, based on optimal estimation theory, for the integration and adaptation processes. This model is tested empirically in the subsequent sections. Section 2.4 provides an overview of the experiments. Section 2.5 describes the baseline experiment examining localization of visual, auditory, and visuo-auditory stimuli in the azimuth. Section 2.6 examines adaptation of visual and auditory maps to an experimentally-induced displacement in the normal visuo-auditory relationship. Section 2.7 examines adaptation to added variability (zero-mean, constant variance noise) in the visuo-auditory relationship. Section 2.8 describes how adaptation to an induced visuo-auditory displacement at one point generalizes to other locations in the azimuth. Section 2.9 describes the control experiments. Finally, Section 2.10 summarizes the results of the experiments in the context of the optimal estimation model.

[1] A spatial *map* is defined as a topographic arrangement of cells whose receptive fields are related in an orderly manner to locations in space.

2.2 Background

The integration of sensory modalities has been studied extensively within both psychology and neuroscience. This section reviews the relevant background literature from both psychophysical and neuroscientific approaches to the study of multisensory integration. Since the psychophysical experiments in this chapter focus on the integration of spatial information from auditory and visual modalities, a brief review of the psychophysics of auditory and visual localization will also be provided.

2.2.1 Psychophysics

Auditory Localization

The ability to correctly orient to auditory stimuli is present in humans within the first 10 minutes after birth (Wertheimer, 1961). There are two types of cues, monaural and binaural, upon which this ability depends. Monaural cues arise from the sound filtering properties of the pinna (the outer ear) and from head movements. For normally hearing listeners, monaural cues act primarily to resolve front-back ambiguities (for reviews of auditory localization in humans see Scharf & Houtsma, 1986 and Blauert, 1983). By far the most important cues for auditory localization, especially in the horizontal plane, are binaural. These can be divided into two classes: those arising from interaural time differences (ITD), and from interaural intensity differences (IID).

Interaural time differences arise both from the fine structure of an acoustic signal, in the form of phase differences, and from the coarse structure, in the form of differences in arrival time or acoustic signal envelope (defined as the amplitude modulation of the waveform). Rough calculations based on the average path between the ears (about 23 cm) and the velocity of sound (344 m/s) reveal that for signals above 750 Hz arising from one side, phase difference cues are ambiguous between lead and lag (Scharf and Houtsma, 1986). This ambiguity sets in at higher frequencies closer to midline. Phase difference cues are therefore not reliable at high frequencies. On the other hand, the cues based on arrival time do not present such ambiguities.

Interaural intensity differences arise from the filtering properties of the head, which can cause level differences between the ears of up to 40 dB (i.e. a hundred-fold intensity difference; Blauert, 1983). Since the head acts essentially as a low-pass filter, IID cues are most effective at higher frequencies. Interestingly, IID provides the best cues around 0° (straight-ahead) and 180° (Fedderson et al., 1957). This is because, even though there is no IID at 0°, the rate of change in IID per degree is highest around 0°.

To a first approximation, the psychophysics of auditory localization are well characterized by the "duality theory," which states that frequencies below 2000 Hz are localized based on ITD cues and frequencies above 4000 Hz are localized based on IID (Rayleigh, 1907). The first clear quantitative support for this theory was given by Stevens and Newman (1936) who examined localization errors as a function of

frequency, and found a peak in errors between 2000 Hz and 4000 Hz, with relatively good performance below and above this range. This suggests that neither binaural cue works well in this transition region. Stevens and Newman also showed that localization was best for broad-band signals, such as clicks (with an average error of 8°) and hisses (5.6°).

Mills (1958) used a different measure, the minimum audible angle (MAA), to study the precision of localization. The MAA is defined as the smallest angle of displacement of a sound source needed to tell whether the sound has moved left or right. Varying the stimulus frequency from 250 Hz to 10 kHz and the location from 0° to 90° in the azimuth in 15° intervals, Mills found that the precision of localization was poorest between 1500 Hz to 2200 Hz and above 5000 Hz. Localization was best at 0° for all frequencies with an MAA uniformly below 4° , and became monotonically worse away from straight-ahead.

Mills' findings have been confirmed since, and it is now well established that the sizes of errors and the response variability are smallest directly in front and increase towards the periphery. For example, Middlebrooks and Green (1991) report broad-band stimulus localization errors of 2° to 3.5° directly in front in the azimuth, increasing to as much as 20° in some rear locations. Once a sound is audible its level only marginally improves localization (Scharf and Houtsma, 1986).

Visual Localization

Several factors contribute to the ability to visually localize a stimulus. Clearly, the principal cue for visual localization is the retinal coordinate of the stimulus. The spatial capabilities of the visual system vary over the retina and across lighting conditions. The fundamental measure of the spatial capabilities is grating acuity, measured by testing whether subjects can tell whether a small grated patch has vertically or horizontally oriented bars. Young adults exhibit an average grating acuity of 42 cycles/degree (Wilson et al., 1990). Acuity is highest at the fovea, at about 60 cycles/degree (cpd) and falls off to about 5 cpd at 30° eccentricity from the fovea. Visual acuity also varies with lighting levels, with a distinction between the photopic system dominated by input from the cones concentrated in the central region of the retina, and the scotopic system, dominated by input from the rods in the periphery (Sekuler and Blake, 1990). The photopic system operates in higher light conditions and has high resolution and low sensitivity (the ability to detect small amounts of light); in contrast, the scotopic system operates in dim light conditions and has low resolution and high sensitivity.

Unfortunately, though acuity may define the limits of the system it is hard to relate it to a more naturally defined perceptual or perceptuomotor notion of localization. A more direct measure of visuomotor localization is the accuracy of saccadic eye movements to targets at various eccentricities. Voluntary human saccades can range in size from 3 min arc to 90° (Robinson, 1987). Primary saccades normally fall short of their target by about 10%; this appears to be a deliberate strategy of the saccadic

system whose purpose is not known (Becker & Fuchs, 1969; Henson, 1978, as reviewed in Robinson, 1987). The primary saccade is usually followed by a corrective saccade that puts the eye on target.

Visual localization can also be measured through pointing movements. To point to a target, its retinal coordinates must be integrated with information on the eye position in the head, and head position relative to the body, converting the location of the target into body-centered coordinates appropriate for movement (Matin, 1986; Andersen, 1987). These additional coordinate transformations undoubtedly add some biases and variability to the measure of visual localization.

Multisensory Integration

The extensive psychophysical literature on intersensory interactions (reviewed, for example, in Welch & Warren, 1986) reveals that the perception of a sensory input is often modulated by the inputs to a number of other modalities. These interactions can often be mediated through secondary causes. For example, vestibular inputs can affect the perception of the location of a sound (Clark and Graybiel, 1949; Graybiel and Niven, 1951; Lackner, 1974b; Lackner, 1974a) and the orientation of a visually displayed line (Day and Wade, 1966) by altering the perceived orientation of the body with respect to gravity. We will primarily review the integration of visual and auditory stimuli, focusing on two classic phenomena, the "McGurk effect" (McGurk and MacDonald, 1976) and the "ventriloquism effect" (Howard and Templeton, 1966), which shed light on the visuo-auditory integration in speech perception and spatial localization, respectively.

The "McGurk effect." The "McGurk effect" was the first clear demonstration of the influence of vision on speech perception (McGurk and MacDonald, 1976). Subjects listened to spoken syllables while viewing the image of a person speaking a different syllable, and asked to report the identity of the syllable heard. The perceived syllable was neither the one presented visually nor the auditory one, but intermediate to the two. For example, when listening to a syllable which in isolation they would perceive as "ba" while viewing a mouth producing the syllable "ga," subjects would perceive the syllable "da." This finding has been confirmed and extended in many ways since. For example, it has been shown that the McGurk effect can be modulated by the amount of noise in each channel (Sekiyama and Tohkura, 1991). Nuclear magnetic resonance imaging studies have also begun to probe the neural substrate for this effect, showing that the sight of the lips modifies activity in the auditory cortex (Sams et al., 1991).

Braida (1991) reviews the principal computational models of integration that could account for visuo-auditory (e.g. McGurk) and related auditory-tactile effects in speech. Two of the models reviewed are based on optimal integration from a statistical perspective, while the third is based on Fuzzy Logic. In the "pre-labelling" model, the raw sensory data is combined across modalities before being categorized using a

multidimensional classification algorithm; while in "post-labelling" and the Fuzzy Logic model, the inputs are first categorized by each modality and then integrated using Bayes' rule. Using syllable confusion matrices obtained from several studies, Braida convincingly argues that while the post-labelling and Fuzzy Logic model cannot easily account for the data, the pre-labelling model fits the data well. Braida's results for speech perception suggest that the approach of predicting multisensory performance from unisensory data and a statistically-based optimal processing model is fruitful.

The "ventriloquism effect." The "ventriloquism effect," a term coined by Howard and Templeton (1966)[2], is perhaps the most commonly studied spatial illusion arising from the integration of visual and auditory stimuli. As suggested by the name, the effect arises when the perceived location of a sound shifts in the direction of a concurrent visual stimulus. As with many illusions, this effect reflects some basic properties of the perceptual system; in the case of ventriloquism this property is that vision dominates in our perception of space.

To merge visual and auditory stimuli in space, two criteria seem to be essential: the stimuli should be approximately synchronous and their locations not too distant (Bregman, 1990). Visual and auditory stimuli up to 30° apart can be merged under conditions of good synchrony (Jack and Thurlow, 1973). When the location of the sound is not fully captured by the visual stimulus, the sound is typically perceived to be between the true and visual location. The perceived location of the visual stimulus, however, is rarely altered by the sound.

A quantitative comparison of visual-auditory, visual-proprioceptive, and auditory-proprioceptive interactions yields an interesting pattern of intersensory biases (Welch and Warren, 1986). The visual bias on perceived proprioceptive location reported by Hay et al. (1965) and subsequent studies ranged from 60 to 75 % of the discrepancy, while the effect of proprioception on vision ranged from 16 to 40 %. A 60 % bias of vision on proprioception signifies that if vision and proprioception are relatively displaced (e.g. using prisms), the proprioceptively perceived location of a limb will be shifted by 60% of the displacement towards the visual location. In nearly all of these studies the sum of these two effects was not statistically different from 100 %. When vision and audition are discrepant the sound is heard near or at its seen location (Welch and Warren, 1986; Jack and Thurlow, 1973; Stratton, 1897a; Pick et al., 1969). This effect of vision on audition amounts to 40 to 80 % of the discrepancy. Attempts to find a biasing effect of audition on vision have had little or no success (Pick et al., 1969; Warren and Pick, 1970). Proprioceptive bias on audition ranged from 50 to 80 %, whereas auditory bias on proprioception ranged from 1 to 18 % (Fisher, 1968; Pick et al., 1969; Warren and Pick, 1970).

The results for intersensory bias correlate strongly with the localization acuity

[2]This effect has also been called "visual capture" (Hay et al., 1965).

of the different senses. Fisher (1960) compared the acuity of vision, audition, and touch/proprioception by having subjects reach, without feedback, to targets within each of these modalities. While it should be noted that these acuity measurements are inflated by variability in the reaching response, the order of decreasing acuity he found, vision, proprioception, followed by audition, reflects the order of intersensory biases.

On the whole, the relation between acuity and intersensory bias found in the above studies is consistent with the optimal estimation model of integration proposed in this chapter. The theoretically appealing relation between acuity and intersensory bias has been noted by many researchers in the past (Choe et al., 1975; Fisher, 1968; Howard and Templeton, 1966; Kaufman, 1974), who proposed variants of the "modality precision hypothesis" (reviewed by Welch & Warren, 1986). Under this hypothesis, intersensory discrepancy will always be resolved in favor of the more precise of two modalities. The optimal estimation model is a formalization of the modality precision hypothesis. The model makes mathematically explicit the relation between the acuity, or more precisely, the reliability, of a source and it's effect on the sensory interpretation of another source. This relation is derived from an underlying principle of optimal integration and allows quantitative predictions to be made, for example, on the spatial distribution of adaptation to visuo-auditory displacements.

Visuo-Auditory Adaptation

When a spatial discrepancy is introduced between vision and audition, the auditory modality adapts. Stratton (1897a, 1897b; reviewed by Blauert, 1983 and Welch, 1978), reports in passing that when subjects wore eyeglasses that turned the visual field upside down, auditory events were also inverted as long as they remained in the visual field. Although subjects wore the perturbing eyeglasses for an extended period of time, this effect should not be considered evidence for adaptation but rather for "ventriloquism" as no aftereffects were reported. Evidence for short-lived adaptation, in the form of an aftereffect of visuo-auditory rearrangement, was found by Klemm (1909, 1918) using a setup in which two microphones were placed in front of two sound-generating hammer devices to the left and right of a subject. The signal from the microphones was presented to the opposite ears via headphones and subjects were asked to judge which hammer the sound was emanating from. With the eyes open, the location of the sound was captured by the visual display of the hammer and appeared on the same side as it. Upon closing the eyes, the sound continued to appear to emanate from the side predicted by the visual display—an aftereffect—and only gradually shifted to the opposite side. Studies of visuo-auditory rearrangement continued with Wooster (1923) and Ewert (1930) who, using prismatic displacements of vision, also showed the powerful effect of an object's visible locus on its apparent auditory position.

Held (1955) conducted the first study that systematically showed adaptation to auditory displacements. Using a pseudophone, a device consisting of two small hearing

aids connected via a rotatable rod to miniature earphones in the subject's ears, Held was able to arbitrarily rotate the input into the two ears relative to the head. He found that a 22° displacement was not only completely visually captured, but after several hours of exposure while moving actively, induced a 10° shift in the perceived auditory midline (i.e. the direction of sounds which subjects judge to be straight ahead). From this and subsequent studies, Held concluded that auditory adaptation results from association between interaural time differences and movement of the body or head.

Many studies have since investigated the effects of visual and auditory rearrangement on auditory localization (e.g. Canon, 1970; Lackner, 1973, 1974a; Radeau & Bertelson, 1974; Shinn-Cunningham, 1994). Like Held's studies, these have found that adaptation seems to be facilitated by active movement, although it can also occur in its absence (Canon, 1970; Radeau and Bertelson, 1974). Summarizing Welch (1978), the basic effects of auditory-visual rearrangement can be attributed to three possible sources: (1) recalibration of interaural difference cues, (2) shift in the felt position of the head, and (3) specific changes in auditory-motor coordination.

2.2.2 Neuroscience

The integration of information from multiple sensory modalities and the plasticity in the relationship between the senses pose interesting problems for neuroscience. Multisensory integration phenomena, such as visual capture and the McGurk effect, suggest that information from multiple modalities, which arrives to distinct areas of the brain in very different representations, eventually converges at some common locus in a common representation. Adaptation experiments imply that discrepancies between the senses and between the predicted and sensed outcome of movements can cause rapid plastic changes in the functional organization of the nervous system. I will first selectively review evidence for the convergence of visual, auditory and proprioceptive information in the CNS. I then turn to one of best studied multisensory and sensorimotor areas of the CNS: the superior colliculus. This area is especially relevant to all the studies in this chapter as it is thought to be principal site of visual, auditory and somatosensory convergence that mediates orienting movements.

Multisensory Neurons

Neurons which respond to inputs from more than one sense can be found in multiple areas of the brain. Some areas, such as the reticular formation which plays a primary role in arousal, receive multisensory inputs of a nonspecific nature. In other areas the neural responses to inputs from different senses are related in a precise manner. For example, in higher visual cortex (i.e. V4) some neurons are tuned to specific line orientations whether they are presented as visible bars or as bars that are felt by the hand but not seen (Maunsell et al., 1989). Similarly, visual cells in parastriate cortex (visual association areas 18 and 19) of the cat respond to acoustical stimulation, with

auditory receptive fields organized in a systematic way relative to the visual receptive fields (Morrell, 1972). This correspondence is not, however, one-to-one: while the visual receptive fields were localized in both vertical and horizontal dimensions, the acoustical receptive fields were localized in the horizontal dimension and elongated in the vertical dimension. Cells sensitive for moving stimuli had the same direction selectivity in both modalities.

We will not provide a review of the literature on multisensory neurons but refer the reader to Stein & Meredith (1993). Summarizing their review, neurons receiving multisensory inputs have been found throughout the cortex, basal ganglia, various regions of the cerebellum, some nuclei of the thalamus, and the superior colliculus.

The Superior Colliculus

The superior colliculus, and its non-mammalian homologue, the optic tectum, is a midbrain structure involved in attentive and orientation behavior (Stein and Meredith, 1993; Kandel et al., 1991). The superior colliculus (SC) is composed of seven layers of cells, operationally divided into two parts: superficial (layers I-III) and deep (layers IV-VII). The superficial layers receive visual inputs both directly from the retina and from visual cortex. The deep layers receive visual, somatosensory, auditory and motor-related inputs (Wickelgren, 1971; Harris et al., 1980; Stein and 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 and Sterling, 1969). The outputs of the superior colliculus project to brain stem and spinal cord areas directly involved in positioning the peripheral sense organs. Though commonly considered part of the eye movement control system, the SC in fact also plays a primary role in orienting movements of the head, limbs and, in species that can move them, ears and whiskers (Harris et al., 1980; Sparks and Nelson, 1987; DuLac and Knudsen, 1990; Guitton and Munoz, 1991; Stein and Meredith, 1993).

A fundamental problem faced by the superior colliculus is that while auditory information is represented in head-centered coordinates, visual information is represented in retinal coordinates. In order to maintain visual and auditory maps in register as the eyes move in orbit, one of three things must occur: either (1) the visual receptive fields dynamically reorganize to match the auditory map, (2) the auditory receptive fields dynamically reorganize to match the visual map or (3) one or the other system is shut down to prevent conflict (Pöppel, 1973). Harris, Blakemore, and Donaghy (1980) found that in cats this problems is circumvented though a behavioral strategy: every eye saccade is followed by a head movements so as to maintain the eyes centered in orbit. Visual and auditory maps are therefore only momentarily out of register. However, primates, including humans, often maintain their eyes fixated on peripheral targets, and therefore the registration problem cannot solved through this same behavioral strategy. In monkeys, Jay & Sparks (1984) found

that auditory receptive fields shifted with changes in eye position, allowing the maps to remain in register. This raises the interesting unanswered question of how this on-line dynamic reorganization takes place.

Stein et al. (1989) studied visuo-auditory integration in the superior colliculus using a behavioral paradigm modeled after neurophysiological experiments for recording from collicular neurons. Cats were required to fixate directly ahead and orient to visual and auditory stimuli in one of three conditions. In the spatially-coincident condition, simultaneous visual and auditory stimuli of varying intensities were presented at random locations, but with no discrepancy between the visual and auditory location. They found that combining stimuli enhanced the probability of a correct response significantly more than would be predicted by responses to unimodal stimuli, especially at peripheral locations where both modalities were less accurate. In the spatially-disparate condition, animals were trained to orient to visual stimuli while ignoring auditory stimuli, and then tested with simultaneous visual and auditory stimuli that were relatively displaced by 60° . This condition resulted in a significant increase in errors in localizing the visual stimuli. On many trials the animals moved directly to a position halfway between visual and auditory stimuli, possibly indicating the locus of an integrated signal. The third, spatial resolution, condition was similar to the spatially-disparate condition except that the visuo-auditory displacement was varied randomly during testing. A systematic pattern of effects emerged. The auditory stimulus facilitated visual localization only when it was displaced more laterally (peripherally) than the visual stimulus, and inhibited visual localization when it was more medial.

These behavioral results are, on the whole, consistent with the neurophysiological data (Stein and Meredith, 1993). As predicted by the spatially registered receptive fields of multisensory neurons, spatially coincident stimuli produce enhancements and spatially disparate stimuli produce either depression or no effect. The results of the spatial resolution condition can be explained by considering two facts: (1) the visual targets at which the effect was observed were located at $\pm 30°$ from center and (2) auditory receptive fields in these lateral areas can be quite large, extended from 20° to 120° into the periphery. Therefore, any auditory stimulus lateral to the target would enhance the visual activity of a large portion of the population of neurons encoding the correct location, accounting for the enhancement observed.

Knudsen and colleagues have extensively studied adaptation to visuo-motor and visuo-auditory displacements and their effects on the neural representations of space in the optic tectum of the barn owl. Knudsen and Knudsen (1989a,1989b) showed that prismatically-induced displacements of visual space imposed from birth, while barely modifying visual localization, induced significant adaptation of auditory localization. This suggests that, in contrast to primates, owls have a relatively hard-wired representation of visual space in the optic tectum. Furthermore, visual inputs, even when incorrect (in the sense that they lead to consistent motor errors), seem to recalibrate the representation of auditory space. In blind-reared owls, the maps of

auditory space in the optic tectum developed abnormally, with erratic progressions in the azimuth of receptive fields, and erratic, stretched or upside-down representations of elevation (Knudsen et al., 1991). This again suggests 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 and Knudsen, 1993). Further research needs to be done to determine the signal driving adaptation in the inferior colliculus (c.f. the model proposed by Pouget, Deffayet & Sejnowski, 1995). This and many other questions on the neural basis of visuo-auditory adaptation are current topics of research.

2.3 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. Clearly, these two problems need not be solved sequentially, or by separate neural processes.[3] The first problem is the coordinate transformation problem and is the topic of Part II of this thesis. In this section we focus on the second problem, the integration problem, assuming that the coordinate transformation problem has already been solved.

2.3.1 Integration

Consider n signals originating from separate modalities which have already been converted into a common representation. The statistical estimation framework assumes that each of these signals is a noisy measurement of some underlying quantity that is to be estimated, such as the location of an object. Each measurement, x_i, $i = \{1 \ldots n\}$, results from a common underlying signal x^* corrupted by additive noise ϵ_i:

$$x_i = x^* + \epsilon_i. \tag{2.1}$$

The goal is to estimate x^* optimally from the measurements. Optimality, defined in the statistical sense of maximizing the likelihood of the estimate given the data, depends on the assumptions about the noise ϵ_i. Two cases can be distinguished depending on the nature of this noise.

[3] For example, Braida's (1991) pre-labelling model for the integration of speech signals does not solve these problems separately.

Independent noise

If the noise in each signal is independent from all the other noise sources, the likelihood of the measurements given an estimate x^* can be factored:

$$L(x_1, \ldots, x_n; x^*) = \prod_{i=1}^{n} p_i(x_i; x^*), \tag{2.2}$$

where $p_i(x_i; x^*)$ defines the statistical model for the noise process corrupting measurement i. To obtain the maximum likelihood estimate (MLE) of x^* it is often easier to maximize the log of (2.2). Denoting the MLE of x^* by $\hat{x}$, we obtain

$$\hat{x} = \arg\max_x \sum_{i=1}^{n} \log p_i(x_i; x). \tag{2.3}$$

This is the general form of the maximum likelihood integration rule under an independent noise assumption.

We focus on the model in which each noise source has a zero-mean Gaussian distribution of differing variance σ_i^2,[4] denoted by

$$p_i(x_i; x) = N(x, \sigma_i^2) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\{-(x_i - x)^2/2\sigma_i^2\}. \tag{2.4}$$

The log likelihood rule (2.3) becomes

$$\hat{x} = \arg\min_x \sum_{i=1}^{n} \frac{(x_i - x)^2}{2\sigma_i^2} + c \tag{2.5}$$

where c is a constant independent of x and can therefore be ignored. The maximum of (2.5), which can be obtained by setting its derivative with respect to x equal to 0, is

$$\hat{x} = \sum_{i=1}^{n} \frac{\sigma_i^{-2} x_i}{\sum_{j=1}^{n} \sigma_j^{-2}} = \sum_{i=1}^{n} w_i x_i, \tag{2.6}$$

where $w_i = \sigma_i^{-2} / \sum_{j=1}^{n} \sigma_j^{-2}$. This rule states that the optimal estimate linearly combines the signals, weighted by their inverse variances. The variance of this estimate

[4] Or in the multivariate case, covariance matrix Σ_i. The univariate case will be presented throughout, though the multivariate extension is straightforward.

is

$$\sigma_{\hat{x}}^2 = (\sum_{i=1}^{n} \sigma_i^{-2})^{-1}, \tag{2.7}$$

which is smaller than the variance of each of the signals and of any other unbiased estimator. We therefore refer to the estimator given by (2.6) as the *minimum variance estimator* (MVE).

Non-independent noise

Factorization of the likelihood is not in general possible if the noise sources are not independent. However, a special case of the non-independent noise problem, correlated Gaussian noise, is interesting and tractable. Define $\mathbf{x}$ to be the vector $[x_1, \ldots, x_n]$ of measurements with covariance matrix V. The estimate is a linear combination of the measurements,

$$\hat{x} = \mathbf{w}^T \mathbf{x}, \tag{2.8}$$

where $\mathbf{w}$ is the vector of weightings. The minimum variance estimator can be obtained by minimizing a cost function consisting of the variance of $\hat{x}$ and a Lagrange multiplier for the constraint that the weights sum to 1:

$$C = \mathbf{w}^T V \mathbf{w} + \lambda(\mathbf{w}^T \mathbf{1} - 1), \tag{2.9}$$

where $\mathbf{1}$ is an n-dimensional vector of ones and λ is the Lagrange multiplier. The minimum of C is obtained when

$$\mathbf{w} = \frac{V^{-1}\mathbf{1}}{\mathbf{1}^T V^{-1} \mathbf{1}}. \tag{2.10}$$

For example, for two Gaussian sources, x_1 and x_2, with variances, σ_1^2 and σ_2^2 and covariance σ_{12}, the MVE is

$$\hat{x} = \frac{(\sigma_2^2 - \sigma_{12})x_1 + (\sigma_1^2 - \sigma_{12})x_2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}. \tag{2.11}$$

Temporal integration

Until now we have examined integration across different modalities ignoring the temporal nature of many sensorimotor integration problems. The optimal estimation

framework extends in a straightforward manner to integration over time.[5]

Consider a single sensor receiving a sequence of measurements x_t and maintaining an estimate $\hat{x}_t$ at time t. The minimum variance update rule for $\hat{x}_t$ can be derived from (2.6) simply by considering the previous estimate as another measurement:

$$\hat{x}_t = \frac{\sigma_{x_t}^{-2} x_t + \sigma_{\hat{x}_{t-1}}^{-2} \hat{x}_{t-1}}{\sigma_{x_t}^{-2} + \sigma_{\hat{x}_{t-1}}^{-2}}. \tag{2.12}$$

The variance of the estimate follows the recursion

$$\sigma_{\hat{x}_t}^{-2} = \sigma_{\hat{x}_{t-1}}^{-2} \frac{\sigma_{x_t}^{-2}}{\sigma_{\hat{x}_{t-1}}^{-2} + \sigma_{x_t}^{-2}}.$$

For example, for a sequence of equal variance inputs we obtain an integration rule of the form:

$$\hat{x}_t = \frac{1}{t+1} x_t + \frac{t}{t+1} \hat{x}_{t-1}, \tag{2.13}$$

with variance converging to zero at a rate of $1/t$.

The Kalman Filter. A particularly useful and general form of estimator resulting from the minimum variance fusion principle is the Kalman filter (Kalman and 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:

$$x_{t+1}^* = A x_t^* + B u_t + v_t, \tag{2.14}$$

where u_t is some input or control signal that we can observe and v_t is zero mean noise. Second, the measurements observed, denoted by y, are related to the state through another linear equation:

$$y_t = C x_t^* + w_t, \tag{2.15}$$

where w_t is again zero mean noise. The basic idea of the Kalman filter is that an optimal estimate of the state, $\hat{x}_{t+1}$, can be obtained by fusing the input u_t, the observations y_t, and the previous state estimate $\hat{x}_t$ using a model of the dynamical system. Based solely on the previous state, that is, before having observed y_t, the best estimate of $\hat{x}_{t+1}$ is clearly given by $A\hat{x}_t + Bu_t$. Upon observing y_t this estimate is

[5] In engineering, estimation from several static sources is sometimes referred to as *sensor fusion*, while dynamic (i.e. temporal) estimation is referred to as *filtering*.

corrected via a term proportional to the error in the predicted observation, resulting in the following update rule:

$$\hat{x}_{t+1} = A\hat{x}_t + Bu_t + K_t[y_t - C\hat{x}_t].$$

The matrix K_t is the Kalman gain, which weights the previous state estimate and the new input in proportion to their inverse variances. More specifically, if the variance of v_t is Q and the variance of w_t is R, then the Kalman gain is

$$K_t = [A\Sigma_t C^T][C\Sigma_t C^T + R]^{-1}$$

where Σ_t is the covariance of the state estimate. This covariance, in turn, satisfies the following recursion (known as the Ricatti difference equation):

$$\Sigma_{t+1} = A\Sigma_t A^T + Q - K_t[C\Sigma_t C^T + R]K_t^T.$$

Note that substituting $A = C = I$, $B = Q = 0$, $R = \sigma^2$, and $y = x_1$, we obtain (2.13).

The optimality of Kalman filters can be stated in two ways. If the noise is Gaussian, the filter provides the maximum likelihood (minimum variance) 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 (e.g. 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 (e.g. Jordan & Rumelhart, 1992). Using the Kalman filter to model the propagation of state estimation errors during movement, it is possible to address the existence and use of an internal model by the CNS. This is the topic of Chapter 3.

2.3.2 Adaptation

When several sensory sources that are being integrated consistently previde disagreeing information, it is possible that one of them is miscalibrated. The optimal strategy for the nervous system in this case may be to adapt the interpretation of one of the sources or to change the relative weightings of the sources. In this section we derive a learning rule for adaptation from the optimal estimation framework. This 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

$$w_2 = 1 - w_1 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}.$$

If the two signals consistently disagree, say by a constant offset or bias, how much should each modality adapt to incorporate this bias? The simplest supervised learning rule, known alternately as the delta rule, the Widrow-Hoff rule, or the the LMS rule, and derivable from the maximum likelihood framework using a Gaussian noise assumption (Widrow and Hoff, 1960; Rumelhart and McClelland, 1986; Hertz et al., 1991), states that if a true or target value is known, then each input should be adapted in the direction of this target. 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. In our multisensory model there is no explicit teaching signal or target. [6] However, by replacing the target with the minimum variance estimate of x we obtain the following interesting form of the delta rule

$$\begin{aligned} \Delta x_1 &= \eta(\hat{x} - x_1) \\ &= \eta(w_1 x_1 + w_2 x_2 - x_1) \\ &= \eta(w_2 x_2 - (1 - w_1) x_1) \\ &= \eta w_2 (x_2 - x_1) \end{aligned} \tag{2.16}$$

We will call the learning rule given by (2.16) the *weighted delta rule* (WDR). It states that each modality should adapt in the direction of the other in proportion to

[6] The leads to the *problem of veridicality* (R. Held, personal communication). If there is no explicit error signal from the environment how can it be assured that the sensory estimates bear any relation to the quantities being estimated? We assume that although adaptation can occur based purely on sensory discrepancies, the ultimate mechanism that grounds sensory representations to the external world depends on discrepancies between the expected and perceived outcome of movements (Held, 1962).

weighting given the *other* modality. For example, if the two modalities are vision and audition, then it predicts that the auditory map should adapt more in areas of space and under conditions where the visual input is more dominant.

Using t to denote time, the update rule given by (2.16) is

$$x_1^t = x_1^{t-1} + \eta w_2 (x_2^{t-1} - x_1^{t-1})$$

It is easy to show that this rule maintains the minimum variance estimate invariant over time

$$\hat{x}^t = \hat{x}^0 \quad \forall t > 0$$

and that both modalities will eventually converge on this optimal estimate

$$\begin{aligned} x_1^t &\stackrel{t\to\infty}{\longrightarrow} \hat{x}^t = w_1 x_1^0 + w_2 x_2^0, \\ x_2^t &\stackrel{t\to\infty}{\longrightarrow} \hat{x}^t = w_1 x_1^0 + w_2 x_2^0. \end{aligned}$$

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,

$$\Delta x_1 = \eta \sigma_1^2 (x_2 - x_1)$$

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.[7] The variance-weighted delta rule also maintains the minimum variance estimate invariant over time, and converges with both modalities reaching the minimum variance estimate. 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 in Experiment 2 of this chapter, this difference can cause markedly differing predictions regarding the pattern of adaptation.

2.3.3 Related Models

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

[7] Extensions of both the WDR and the VWDR can be derived for the $n > 2$ modality case under very similar assumptions.

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

$$\hat{x} = x_i \quad \text{iff} \quad \sigma_i^2 \leq \sigma_j^2 \quad \forall j. \tag{2.17}$$

Clearly, $\hat{x}$ will have variance σ_i^2, which is generally higher than the variance of the MVE.

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

$$\Delta x_j = \eta(x_i - x_j), \tag{2.18}$$

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 (2.17) predicts that in the presence of a visuo-auditory discrepancy complete visual capture will occur (i.e. ventriloquism). Furthermore, a persistent discrepancy will induce auditory adaptation, but no visual adaptation.

Stochastic integration

A different form of competitive integration occurs if the CNS selects between disprepant 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. For example, if the probability of choosing signal i is inversely proportional to its variance

$$p_i \propto \sigma_i^{-2} \tag{2.19}$$

we obtain

$$\hat{x} = \begin{cases} x_1 & \text{with prob. } p_1 \\ \vdots & \\ x_n & \text{with prob. } p_n. \end{cases} \tag{2.20}$$

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.[8]

[8] In fact, the distribution of this estimator defines a *mixture model* (Titterington et al., 1985), a model commonly used in competitive learning (e.g. Nowlan, 1991; Jacobs, et al. 1991).

The mean of this estimator is the MVE. The variance of this estimator is

$$\sigma_{\hat{x}}^2 = \sum_{i=1}^{n} p_i \sigma_i^2 + \frac{1}{2} \sum_{i,j=1}^{n} p_i p_j (\bar{x}_i - \bar{x}_j)^2, \tag{2.21}$$

where $\bar{x}_i$ denotes the mean of x_i. The second term in (2.21) captures the added variance due to mean discrepancies between the senses. Noting that this term is non-negative and using (2.19) we obtain that the variance is

$$\sigma_{\hat{x}}^2 \geq \sum_i \frac{\sigma_i^{-2}}{\sum_j \sigma_j^{-2}} \sigma_i^2 = n(\sum_j \sigma_j^{-2})^{-1}, \tag{2.22}$$

which is n times larger than the variance of the MVE. A further testable prediction that this rule makes 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, while at each time step it uses a delta rule of the form of (2.18), the stochastic target selection effectively renders it equivalent to (2.16) (the proof follows from taking the expectation of the target). As such, using this rule all the modalities will also converge on the MVE.

2.3.4 Summary

Three computational models of multisensory integration have been proposed. The minimum variance model combines inputs in a statistically optimal way, weighting each by a measure of its reliability. The extension of this model to the dynamic domain is known as the Kalman filter. The competitive model selects the input with the highest reliability while ignoring the other inputs. The stochastic model selects probabilistically between the inputs. Associated with each of these models is a learning rule which can predict the pattern of adaptation resulting from the introduction of an intersensory discrepancy.

2.4 Overview of the Experiments

In the following series of experiments I have sought to establish whether the computational models of integration and adaptation proposed in the previous section can characterize human visuo-auditory localization. All of the proposed models are based

on the principle that multisensory behavior and the pattern and extent of adaptation can be predicted from unisensory behavior. These predictions are quantitative and exact, which makes the models empirically falsifiable.

The common denominator of all the models is their dependence on a measure of reliability, related inversely to the variance in localization. Experiment 1 measures the biases (constant errors) and variances (variable errors) in localization of visual, auditory, and combined visuo-auditory stimuli. Localization is assessed in the plane of the azimuth using a pointing paradigm. Experiment 1 can be considered the baseline from which predictions for all the subsequent experiments will be made. To account for cross-subject variability, subjects in all subsequent experiments also participated in this baseline experiment.

Experiment 2 examines adaptation to a visuo-auditory displacement (an added bias). It is known that the variance in visual and auditory localization changes as a function of location in the azimuth. Based on this variance, each of the computational models predicts a different pattern of adaptation over the azimuth. The models will therefore be tested by comparing these predictions with the actual pattern of adaptation obtained.

Experiment 3 examines adaptation to a zero-mean, randomly varying visuo-auditory displacement (an added variance). Again, as the models specify integration and adaptation rules based on the variance in each modality, it is of interest to examine the effect of artificially changing the variance. If this added variance is interpreted by the sensory system as a change in the reliability of one or the other modality, the relative weightings of the modalities should change.

Experiment 4 examines the pattern of generalization resulting from exposure to a visuo-auditory displacement at a single location in the azimuth. Simultaneous visuo-auditory stimuli are limited to this location and generalization is measured through pointing separately to visual and auditory stimuli across the azimuth. The motivation for this experiment is two-fold. First, like the pattern of adaptation, the pattern of generalization predicted by each model differs. Second, the pattern of generalization can be used to infer properties, such as locality, of the representation of visuo-auditory space. Chapter 4 is dedicated entirely to the issue of inferring the representation of another coordinate transformation—the *visuo-motor* transformation—from its pattern of generalization.

2.5 Experiment 1: Localization of Visual, Auditory, and Visuo-auditory Stimuli

In order to establish the baseline bias and variance of localization we used a sensorimotor paradigm in which subjects pointed to visual, auditory, and visuo-auditory stimuli.

2.5.1 Method

Subjects

Ten right-handed subjects (6 male, 4 female; ages 18-27) participated in this experiment. Subjects were naive to the purpose of the experiment, gave their informed consent, and were paid $7.00 for participation. All subjects had self-reported normal or corrected-to-normal vision and normal hearing.

Apparatus

The experimental setup was designed to achieve three goals: present visual stimuli, present auditory stimuli, and record finger positions (Figure 2-1). Visual stimuli were presented by projecting the Video Graphics Array (VGA) color display of a PC, using a LCD projector (Sayett Media Show), onto a white screen above the experimental table.

Auditory stimuli were presented using a small computer-controlled buzzer (Radio Shack model 273-054; 300-500 Hz buzz, 75 dB sound pressure at 20cm) mounted at the end of a 36 cm rod rotating about the position on the table directly below the subject's chin. The rod was hidden from view by the sound-transparent screen. Two markers were mounted on the end of the rod nearest to the subject so that the rod's angle could be monitored on-line using an Optotrak motion tracking system (described below). The markers were visible to the Optotrak though a small window in the screen; subject's vision of this window was precluded by the chin-rest. A feedback controller, implemented in software on the PC, positioned the rod by controlling a belt-geared DC stepper motor (Superior Electric SLO-SYN model M061; 0.15 deg/step with gearing). The transmission belt also served to reduce audible noise from the discrete stepping.

Finger position was recorded at 200 Hz using an Optotrak 3020 motion tracking system (Northern Digital, Ontario). This was achieved by mounting an infrared light emitting diode (IRED) on the subject's right index finger, the 3D position of which was monitored by the Optotrak to within 0.1 mm. Similar markers were used for the rod. Pointing responses were terminated by the subject by clicking on the button of a PC trackball held in the left hand.

Calibration

Prior to each experiment two forms of calibration were performed. First, the relationship between the two Optotrak markers mounted on the rod and the buzzer's angular position was calibrated. This procedure consisted of marking the approximate center of rotation of the rod and the position of the buzzer at two angular settings of the rod, to the far left and far right. An iterative optimization algorithm then computed, from these marker positions, the best fit (in the least-squared error sense) for the actual center of rotation, rod length, and relation between the two markers mounted

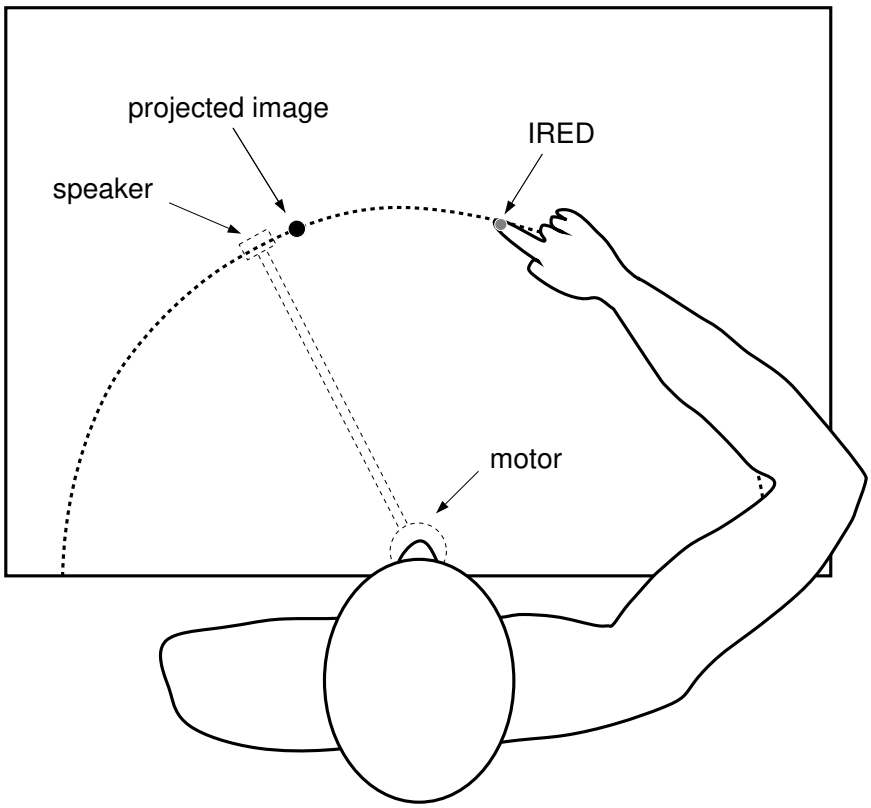

Figure 2-1: Experimental setup. Subjects are seated at a table with an Optotrak marker (IRED) mounted on their right index finger. On the table is a screen visual images are projected. Directly below the screen is a small speaker whose position is controlled by a stepper motor.

on the rod and the buzzer's angular position. Cross-validation tests gave an average calibration error under 0.2° .

The second form of calibration determined the mapping between Cartesian coordinates relative to the table and pixel positions of the projected image. A large grid of sixteen points was projected onto the screen and the actual position of each point was marked in turn using an Optotrak marker. A quadratic regression fit of x and y pixel coordinates to x and y marker position was then performed; the parameters of this fit were used in the experiments to project images accurately onto the plane of the screen. The correlation of the fit was always greater than 0.99 and cross-validation tests gave an average calibration error of less than 2.0 mm.

Paradigm

Subjects were seated at a table with their head resting on a chinrest and an Optotrak marker mounted on their right index finger. The experiment consisted of 12 sessions of 35 trials each, with breaks between each session. Each trial started with a 36 cm radius blue arc projected onto the screen and a 2 cm white fixation cross straight ahead (at 0°) on the arc (Figure 2-2). The cross then disappeared and after a 100 ms delay either a visual (V), auditory (A), or visuo-auditory (VA) stimulus was presented. Visual stimuli were 0.5 cm hollow white squares projected for 100 ms onto the arc; auditory stimuli were 100 ms buzzes originating from below the screen directly underneath the arc; and VA stimuli were simultaneous combinations of V and A stimuli from the same location. Stimuli originated from 35 locations uniformly spaced between $-65°$ and $65°$ in the azimuth. Locations and stimulus modality were completely randomized: each location was tested 12 times and each modality 140 times throughout the experiment.

The subject's task was to point to the location of the stimulus with his or her right index finger. As the subject moved the finger over the screen, a 0.8 cm square cursor spot was projected in the direction that the finger pointed. The cursor spot was always at the same angle as the finger marker with respect to the center of the arc, but was constrained to move along the arc. The purpose of this cursor was two-fold: (1) to prevent fatigue due to the large pointing responses that would be necessary to reach the arc with the finger, and (2) to reduce variability in pointing by providing a cursor for the subject to know the exact location being pointed to. Subjects rapidly became accustomed to pointing with this finger cursor. When the cursor was perceived to point accurately to the stimulus location, the subject pressed the button of a mouse held in the left hand.

Subjects were told that two things were essential: to keep their eyes fixed on the cross whenever it was present, and to try to respond as accurately as possible. It was emphasized that reaction-time did not matter. Subjects were also told that V, A, and VA stimuli would occur randomly and that they should attend to both modalities. Instructions for all conditions were the same: "Point to the stimulus location." The experiment lasted a total of about 50 minutes, and was preceded by a practice session

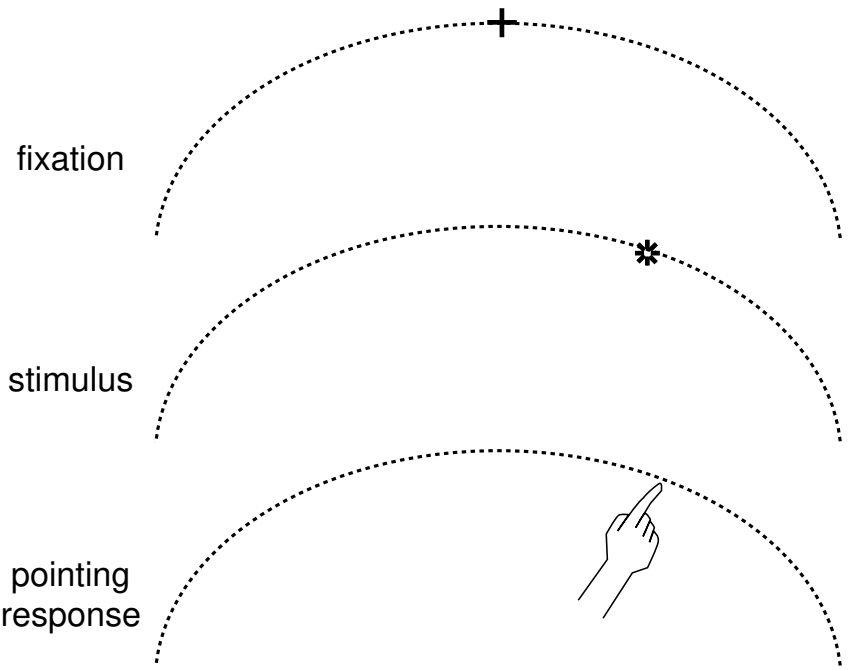

Figure 2-2: Experimental paradigm.

consisting of 18 trials during which the instructions were explained.

Analysis

To assess accuracy of localization the difference between the actual stimulus location and the subject's pointing response, i.e. the localization *error*, was characterized by its mean and variance. For each condition and target location, the mean and standard error of the localization error was averaged across subjects and plotted. This corresponds to bias in localization. Localization variance was computed by subtracting from the error the average bias for that target and stimulus location. These residuals were then squared, and their mean and standard error plotted. For presentation clarity some of the raw data plots were also fitted with 8 degree-of-freedom cubic smoothing splines using the Splus statistical package (Chambers and Hastie, 1992).

2.5.2 Results

Both vision and audition displayed consistent patterns of localization bias (Figure 2-3a & b). For both modalities, bias was not significantly different from zero straight-

ahead, and increased monotonically to the right of straight-ahead.[9] Bias on the left was smaller for vision and displayed an erratic pattern for audition. A similar pattern of auditory bias was found for a different set of 10 subjects in a pilot experiment (not shown). The strong left-right asymmetries in bias displayed for all three stimulus conditions can be mostly accounted for by effects of pointing with the right hand (cf. Left-hand pointing control, section 2.9.2).

The pattern of bias for visuo-auditory stimuli was almost identical to the pattern for visual stimuli (Figure 2-3c). When the three conditions are compared, the visuo-auditory bias is shown to lie uniformly between the visual and auditory bias for stimuli on the right side. Again, the pattern on the left side is more erratic.

The variance of both visual and auditory localization was smallest straight-ahead and increased markedly to the periphery (Figure 2-4a & b). Visuo-auditory localization displayed a pattern of variance almost identical to visual localization (Figure 2-4c). Although visuo-auditory variance was slightly lower that visual variance on the right side, this difference was not statistically significant. Auditory variance in localization was clearly much greater than visual or visuo-auditory variance (Figure 2-4d).

2.5.3 Discussion

The measurement of localization bias and variance in this experiment serves three purposes. First, the results provide a picture of visual and auditory localization which can be compared to the existing literature. Second, by comparing localization of combined visuo-auditory stimuli to localization of separate visual and auditory stimuli, the models of integration presented in this chapter can be tested. Third, the results provide a per-subject baseline from which adaptation can be measured.

The picture of visual and auditory localization provided by these data makes it clear that localization is best straight-ahead for both vision and audition—a finding that is consistent with the existing literature (Mills, 1958; Middlebrooks and Green, 1991). As subjects were fixating on a point straight-ahead, this effect in the visual modality can be attributed to stimulus location on the retina (e.g. the differences between foveal and peripheral acuity could account for the effect). However, other factors, such as the effect of attending to the fixation spot, or the added bias and variance of the pointing response, could also be contributing to the pattern observed. For the auditory modality, the effects of eye position and head orientation are confounded by having subjects fixate straight-ahead. It has been shown that eye position has a significant effect on sound localization (Jones and Kabanoff, 1975; Goldstein and Rosenthal-Veit, 1926 as described in Lackner, 1974). Thus, the pattern observed is most likely a combination of the differential accuracy of localization in the azimuth (e.g. Middlebrooks and Green, 1991) and the effect of fixation straight ahead.

[9] In all plots $-90°$ represents the far left, 0° straight-ahead, and -90° the far right.

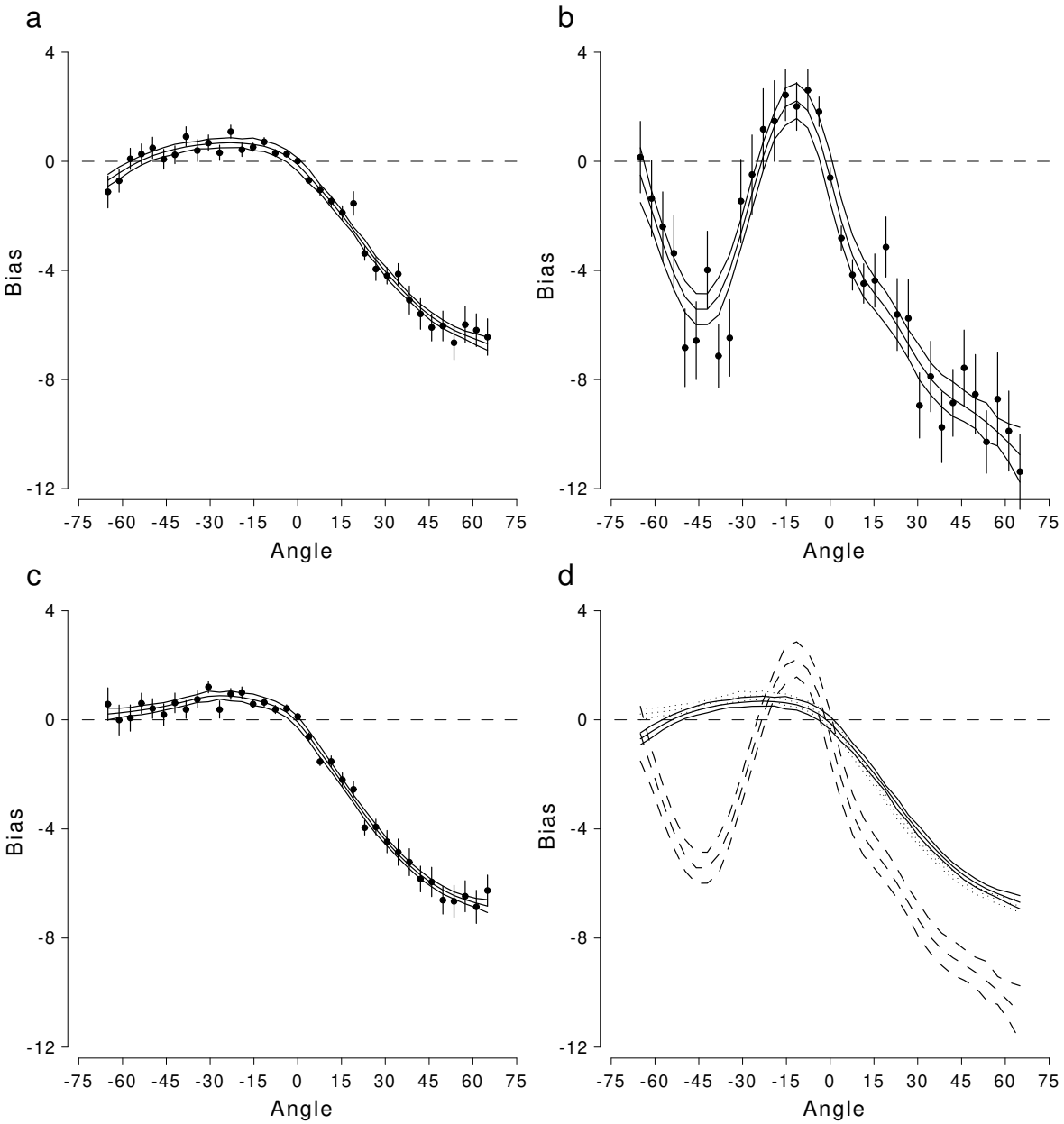

Figure 2-3: Average localization error (bias) plotted as a function of stimulus location (0° is straight ahead). Mean bias (filled circles) is shown with 1 standard error (s.e.) bars and smoothing spline fits (mean $\pm$ 1 s.e. curves), for a) visual, b) auditory and c) visuo-auditory stimuli. d) Comparison of bias in the three conditions: visual (solid), auditory (dashed) and visuo-auditory (dotted).

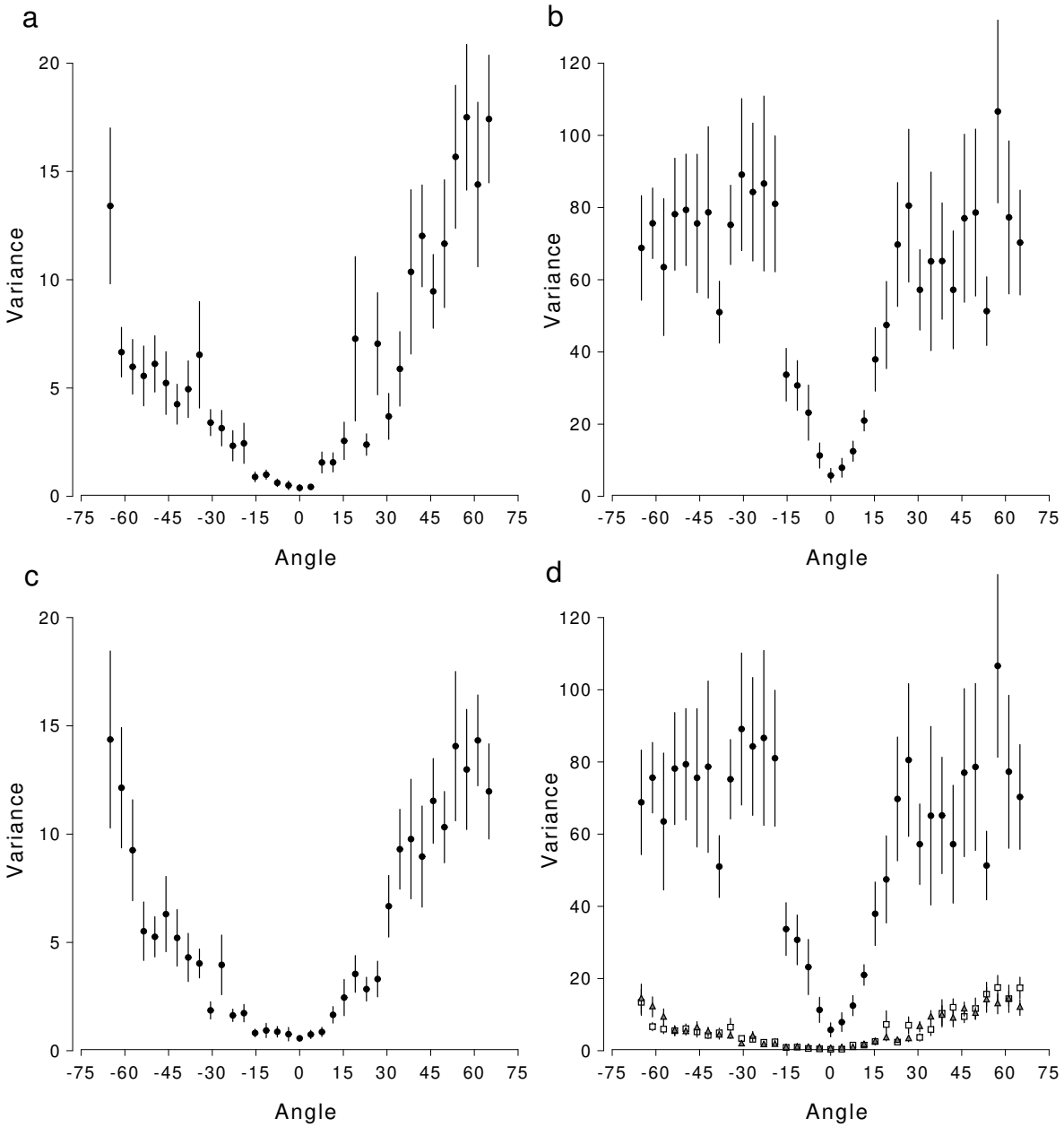

Figure 2-4: Average variance of localization plotted as a function of stimulus location (mean $\pm$ 1 s.e.) for a) visual, b) auditory, and c) visuo-auditory stimuli. d) Comparison the of variance in the three conditions: visual (white squares), auditory (filled circles), and visuo-auditory (gray triangles).

Model predictions

The relative variances of visual and auditory localization suggest that, under any sensible principle of integration, vision will dominate over audition. As discussed in the introduction however, each model makes differing predictions on the amount and pattern of visual dominance. We therefore examined whether the data obtained in this baseline experiment were consistent with the predictions of the three models of integration: minimum variance integration, competitive integration, and stochastic integration.

Minimum variance integration. Under this model, the weighting of each modality used for integration is dependent on localization variance. As localization variance is a function of stimulus location, we used the empirically-observed variances of visual and auditory localization to estimate the optimal weighting function for vision (Figure 2-5). The weights for vision and audition were estimated using (from equation 2.6):

$$w_{\text{vis}} = \frac{\sigma^2_{\text{aud}}}{\sigma^2_{\text{aud}} + \sigma^2_{\text{vis}}} \qquad w_{\text{aud}} = 1 - w_{\text{vis}} = \frac{\sigma^2_{\text{vis}}}{\sigma^2_{\text{aud}} + \sigma^2_{\text{vis}}}. \tag{2.23}$$

Although the variance of auditory localization is smallest around 0° , Figure 2-5 shows that it is relatively larger than the visual variance. Therefore, under minimum variance integration, vision should be most dominant straight ahead. The ratio of variances also suggests that vision is highly dominant overall; the mean proportion for vision is 0.913 ± 0.005. Therefore, the model predicts that (1) visuo-auditory responses will closely resemble visual responses, perhaps with small differences in the periphery, and (2) adaptation will take place mostly in the auditory domain.

Although the second prediction is not addressed by this experiment, the data on the localization of visuo-auditory stimuli address the first prediction. We found that both visuo-auditory bias and variance closely resembled those of visual alone. The small deviations from the visual pattern, especially for locations on the right side, were in the direction predicted by minimum variance integration. That is, the bias was shifted towards the auditory bias and the variance was slightly smaller than the visual variance. The magnitude of these shifts was roughly consistent with the predictions of minimum variance integration; given a 0.9 weighting of vision, minimum variance predicts shifts in bias ranging from 0 to 0.3° and decreases in variances of at most 2 deg^2. All differences vanished around 0° . The pattern on the left side was more erratic. Overall, the finding that visuo-auditory patterns of bias and variance were strikingly similar to visual patterns is consistent with minimum variance integration.

Competitive integration. Given the relative variabilities of visual and auditory localization, the competitive integration model would clearly favor vision over audition. Pointing to visuo-auditory stimuli is therefore predicted to be identical to pointing to visual stimuli. The patterns of visuo-auditory bias and variance found

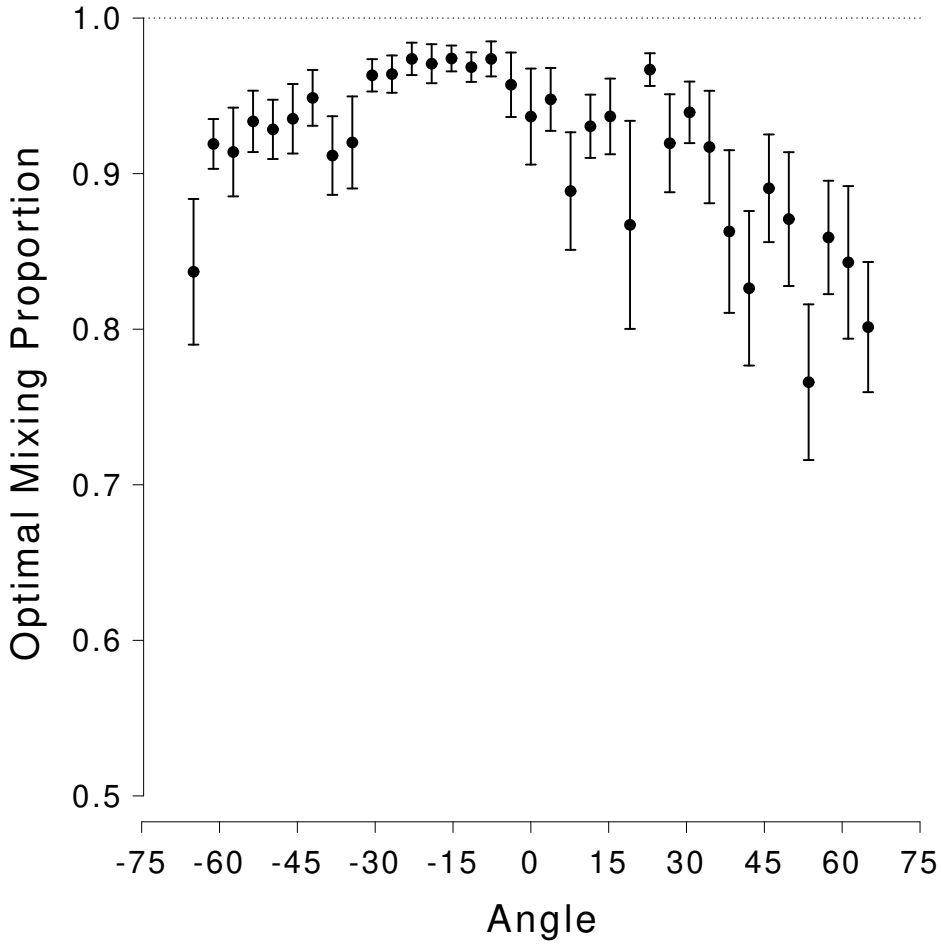

Figure 2-5: Optimal mixing proportion for vision predicted by minimum variance integration. Note that vision dominates the most straight-ahead.

are therefore also consistent with competitive integration.

Stochastic integration. The stochastic model predicts that although visuo-auditory bias will be identical to visual bias, the variance will be substantially larger. Specifically, from equation (2.22) we see that the predicted visuo-auditory variance is about twice the visual variance—a prediction that was not supported by the data.

Although the stochastic model was inconsistent with the pattern of visuo-auditory localization, the almost complete dominance of vision precluded direct testing between the minimum variance and competitive models. Direct testing of these models and their associated adaptation rules is possible if a discrepancy is introduced between vision and audition. In the following experiment we therefore studied adaptation to a discrepancy between vision and audition.

2.6 Experiment 2: Adaptation to a Visuo-Auditory Remapping

Subjects were exposed to either a leftward or rightward shift in the relation between simultaneous visual and auditory stimuli. Adaptation was assessed by measuring any resulting changes in pointing to visual and auditory stimuli.

2.6.1 Method

Subjects

Ten right-handed subjects (5 male, 5 female; ages 18-27) participated in this experiment. Subjects were naive to the purpose of the experiment, gave their informed consent, and were paid $7.00 for participation. All subjects had self-reported normal or corrected-to-normal vision and normal hearing.

Paradigm

Except for the presence of a perturbation, the paradigm was essentially identical to the one used in Experiment 1. The perturbation was absent for the first 3 sessions (trials 1-105; pre-exposure), was introduced gradually, increasing linearly, during the 4^{th} session (trials 106-140), and was present in-full for the last 8 sessions (trials 141-420). As before, only a third of the trials were visuo-auditory; the purely visual and auditory trials throughout the experiment could therefore be used to assess adaptation.

The full perturbation was a 15° displacement between the visual and auditory location of the stimuli. For half the subjects (Group 1) the auditory stimulus was displaced to the left of the visual stimulus; for the other half (Group 2), to the right.

To accommodate these perturbations without extending beyond the limits of the setup, the range for all the visual stimuli was decreased to $-60°$ to $60°$.

Analysis

Adaptation in both visual and auditory localization was analyzed as a function of time and location in the azimuth. To assess the time course of adaptation, the mean bias $\pm$ 1 s.e. for each session was computed by averaging over all target locations and subjects in each group. To assess extent of adaptation as a function of spatial location, the mean bias $\pm$ 1 s.e. for each target was computed by averaging over trials 176–420 and subjects in each group. We will refer to this as the *post-exposure* bias. The choice of trial 176 (the beginning of session 6) as a cut-off was made ad hoc, based on the notion that it would take about 1 session after the onset of full perturbation (session 5) for perceptible adaptation to occur; this choice was maintained in all analyses of spatial adaptation. The same analysis was performed on trials 176–420 of Experiment 1 to establish a baseline measure of bias. The spatial pattern of adaptation was computed by subtracting this baseline bias from the post-exposure bias.

2.6.2 Results

Time course of adaptation

The mean bias of localization for visual, auditory, and visuo-auditory stimuli is shown as a function of trial number in Figure 2-6. While visual localization did not change significantly over the time course of the experiment (Figure 2-6a), auditory localization shifted significantly in the direction opposite the perturbation for both groups (Figure 2-6b). For Group 1 the mean shift (calculated by subtracting the mean pre-exposure bias from the mean post-exposure bias) was $6.8 \pm 0.6°$, while for Group 2 the mean shift was $4.9 \pm 0.6°$. The mean shift combined over both groups was 5.9 $\pm$ $0.4°$, accounting for 39% of the perturbation.

The bias for the visuo-auditory condition, calculated relative to the location of the auditory stimulus, clearly shows the effects of visual capture (Figure 2-6c). When confronted with a $15°$ discrepancy between the location of the visual and auditory stimuli, subjects point to the visual stimulus. For Group 1 the mean shift was $14.2 \pm$ $0.4°$, while for Group 2 the mean shift was $15.2 \pm 0.4°$. Combined over both groups the mean shift was $14.7 \pm 0.3°$, not significantly different from the $15°$ predicted by complete visual capture.

Spatial pattern of adaptation

Adaptation as a function of target location is shown in Figures 2-7 and 2-8 for the auditory and visual modalities, respectively. Auditory localization shifted in the adaptive direction at almost all target locations, although the pattern was variable.

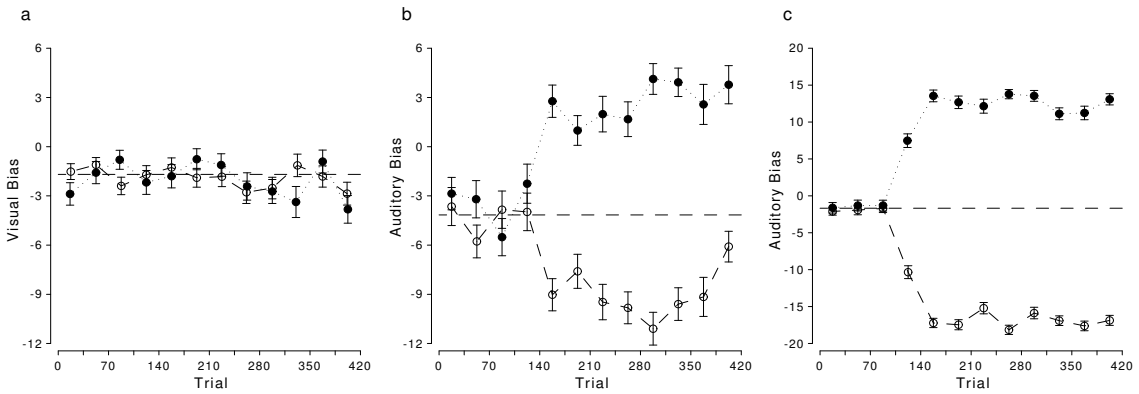

Figure 2-6: Bias as a function of trial number for a) visual, b) auditory and c) visuo-auditory localization. The mean ± 1 s.e. bias for Group 1 (solid circles) and Group 2 (hollow circles) is plotted relative to the baseline bias calculated from trials 1–105 (dashed line). For c) the bias is calculated relative to the location of the auditory stimulus and the shifts therefore correspond to the effect of visual capture and not adaptation (see text).

Visual localization shifted slightly in the direction of greater absolute bias for both groups. As the perturbation was in opposite directions for the two groups, this shift seems unrelated to the perturbation.

2.6.3 Discussion

Introduction of a displacement between the locations of simultaneous visual and auditory stimuli induced a significant shift in auditory localization. That is, when auditory stimuli were presented alone, pointing had shifted towards the location the visual stimulus would have been. These results are consistent with the many previously found accounts of auditory adaptation to visuo-auditory displacements both in humans (Held, 1955; Canon, 1970; Lackner, 1973; Radeau and Bertelson, 1974), and in barn owls (Knudsen and Knudsen, 1989a; Knudsen and Knudsen, 1989b).

In the visuo-auditory condition, subjects pointed 14.7° in the direction of the visual stimulus relative to the auditory stimulus. This was not significantly different from the 15° shift (complete visual capture) predicted by the competitive model of integration. The minimum variance model assuming visual weighting of 0.91 predicts a shift of 13.7° shift, which, although comparable to the shift observed, is significantly different.

This difference may lead one to discard the minimum variance model in favor of the competitive model. However, further evidence from this and subsequent experiments suggests that, contrary to the competitive model, visual and auditory stimuli are indeed combined, although perhaps with a weighting for vision that is greater than

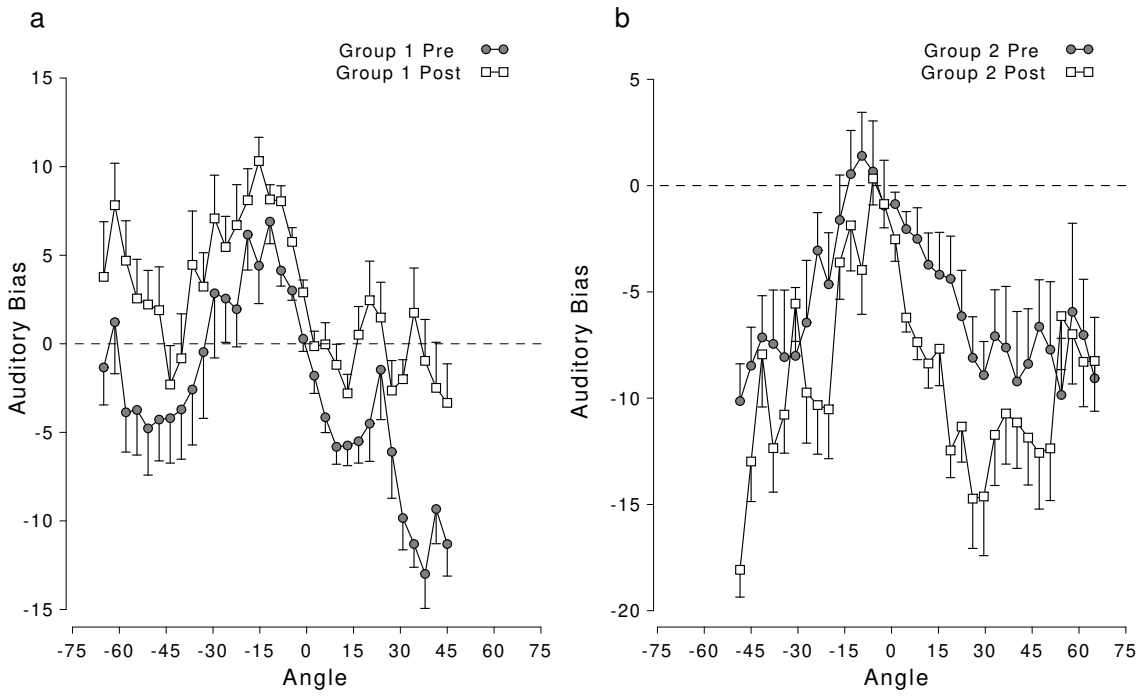

Figure 2-7: Auditory adaptation as a function of target location. a) Group 1 (shifts in the positive direction are adaptive). b) Group 2 (shifts in the negative direction are adaptive).

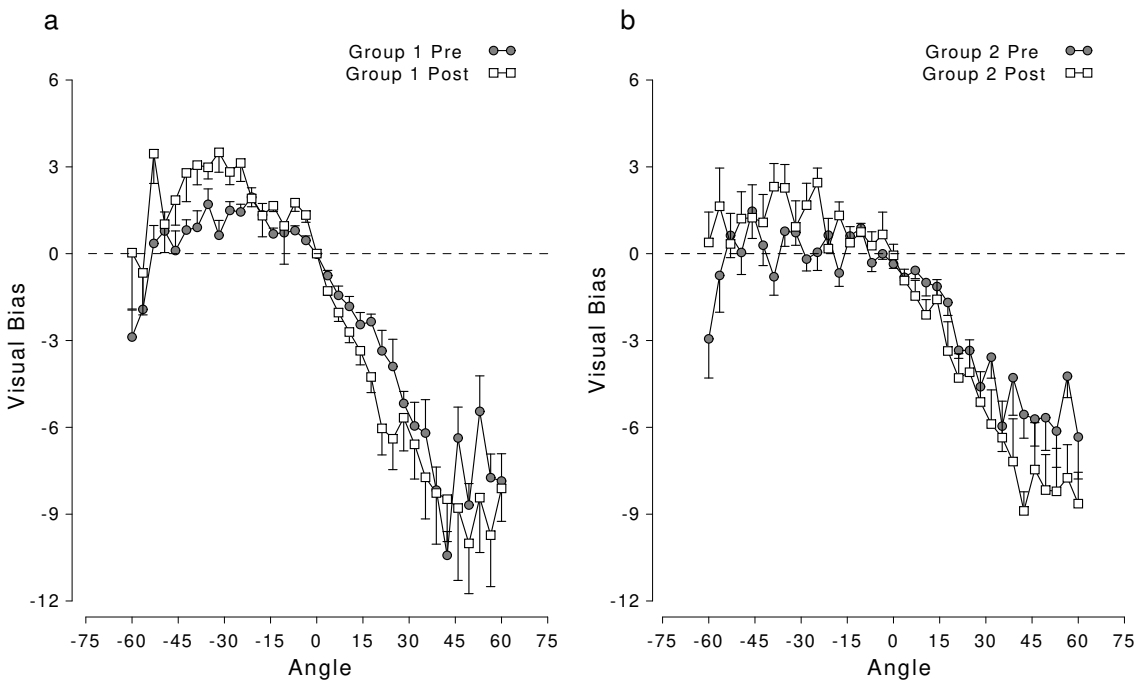

Figure 2-8: Visual adaptation as a function of target location. a) Group 1 (visual shifts in the negative direction are adaptive). b) Group 2 (shifts in the positive direction are adaptive).

0.91. Such an underestimate in the weighting for vision may be due to the fact that variability in pointing was not factored out. That is, variability due to the motor response may inflate both visual and auditory localization variances, decreasing the ratio of auditory to visual variance, and therefore the estimated weighting for vision.

Subjects were completely unaware of the perturbation. After the experiments, subjects were explained the nature of the perturbation and asked whether they had noticed it; none reported having noticed it. Furthermore, as the shifts were measured in the absence of the visual stimulus, they can be considered aftereffects in analogy to the prism adaptation literature (Welch, 1978). Taken together, the presence of significant aftereffects and subjects' unawareness of the perturbation can be taken as evidence that the shifts found were true adaptation and not the result of conscious strategies or "cognitive learning" (Redding and Wallace, 1993; Bedford, 1993).

The computational models of adaptation proposed in the section 2.3 make very explicit predictions on the form of adaptation resulting from a visuo-auditory displacement. We now examine these computational models in light of the observed data.

Delta Rule. The simplest model predicts that adaptation in each modality will be proportional to the displacement introduced:

$$\begin{aligned} \Delta x_{\text{aud}} &= \mu_{\text{aud}}(x_{\text{vis}} - x_{\text{aud}}) \\ \Delta x_{\text{vis}} &= \mu_{\text{vis}}(x_{\text{aud}} - x_{\text{vis}}). \end{aligned}$$

We have introduced two different learning rates, μ_{aud} and μ_{vis}, as it is clear that vision and audition do not adapt equally to a displacement. Specifically, regarding our data on auditory adaptation, this model predicts that adaptation will be equal across all locations of the azimuth. The competitive adaptation rule (2.18) is a variant of this model where the visual modality does not adapt at all ($\mu_{\text{vis}} = 0$). It therefore also predicts that auditory adaptation will be equal across locations in the azimuth.

Weighted Delta Rule. This model, derivable from the principle of minimum variance integration (equation 2.16), predicts adaptation in each modality proportional to the weighting of the other modality:

$$\begin{aligned} \Delta x_{\text{aud}} &= \mu w_{\text{vis}}(x_{\text{vis}} - x_{\text{aud}}) \\ \Delta x_{\text{vis}} &= \mu w_{\text{aud}}(x_{\text{aud}} - x_{\text{vis}}). \end{aligned}$$

In this case, we have collapsed both learning rates into one, μ, as the fact that vision adapts less than audition falls out of the weighting of the two modalities. Specifically, referring to the empirically-derived optimal weighting function (Figure 2-5), the model predicts that (1) audition will adapt most straight-ahead and least in the periphery,

Model	Auditory Adaptation	Visual Adaptation
Delta Rule	equal everywhere	equal everywhere
Weighted Delta Rule	most around 0°	least around 0°
Variance-Weighted Delta Rule	least around 0°	least around 0°

Table 2.1: Summary of model predictions.

(2) conversely, vision will adapt most in the periphery and least straight-ahead, and (3) visual adaptation will be about 10 times smaller than auditory adaptation.

Variance-Weighted Delta Rule. This model, also derivable from the principle of minimum variance integration (equation (2.6)), predicts adaptation proportional to the variance of each modality:

$$\begin{aligned} \Delta x_{\text{aud}} &= \mu\sigma^2_{\text{aud}}(x_{\text{vis}} - x_{\text{aud}}) \\ \Delta x_{\text{vis}} &= \mu\sigma^2_{\text{vis}}(x_{\text{aud}} - x_{\text{vis}}). \end{aligned}$$

Referring to the variances of visual and auditory localization (Figure 2-4), the model predicts that (1) audition will adapt *least* straight-ahead and most in the periphery, (2) similarly, vision will adapt least straight-ahead and most in the periphery, and (3) visual adaptation will be about 10 times smaller than auditory adaptation.

The predictions for all three models are summarized in Table 2.1.

The experimental results are inconsistent with the Delta Rule and the Weighted Delta Rule, and favor the Variance-Weighted Delta Rule. The auditory adaptation plotted as a function of space shows a marked dip at 0° (Figure 2-9). Changes in visual localization, though overall not significant in the adaptive direction, are also most pronounced in the periphery and vanish at 0° (Figure 2-8). The approximate 10:1 ratio of auditory to visual adaptation predicted by the VWDR suggests that vision would adapt by about 4% of the perturbation, or 0.6° . It is therefore likely that, under this model, visual adaptation would have been too small to observe in the data.

Experiment 2 suggests that the pattern of visual and auditory adaptation may be predicted from a simple model, based on optimal estimation principles, which states that each modality adapts in proportion to its variance in localization. A natural question which follows is whether integration and adaptation can be affected by an explicit experimental manipulation of localization variance.

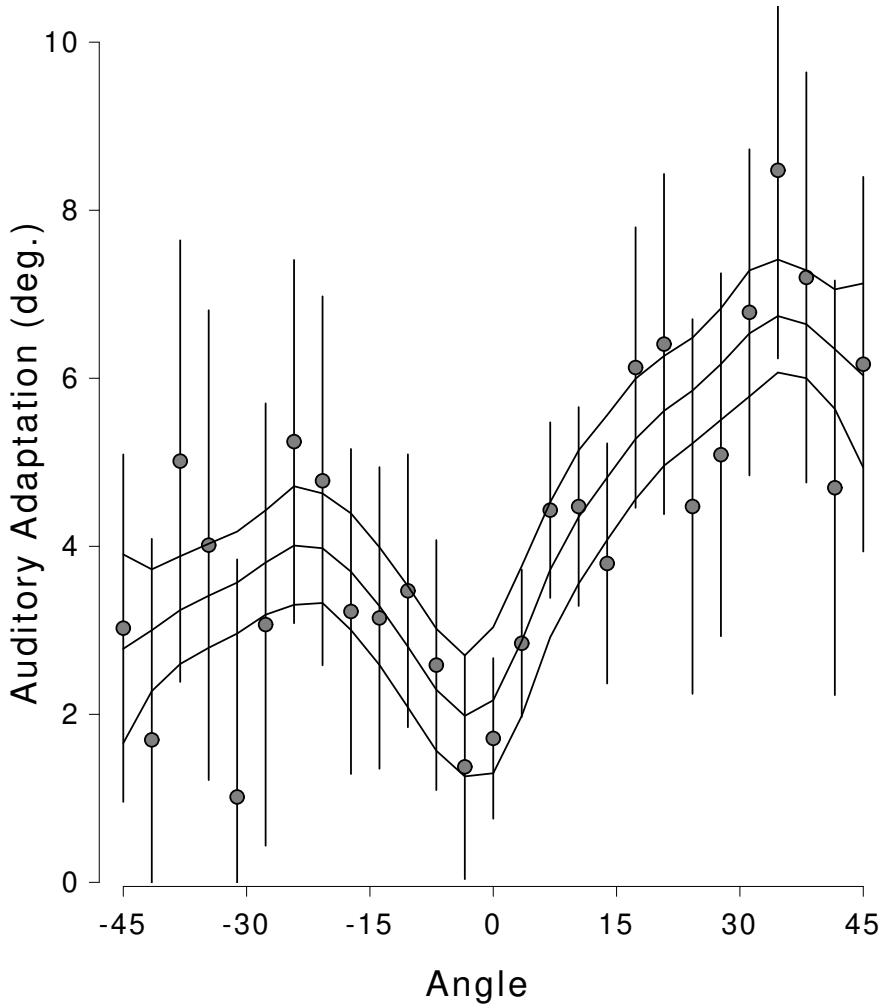

Figure 2-9: Auditory adaptation as a function of location. Data from both groups were combined and means and standard errors plotted. Also shown is an 8 d.o.f. smoothing spline fit with standard error curves.

2.7 Experiment 3: Adaptation to Visuo-Auditory Variance

The computational models proposed in section 2.3 all rely on a statistical measure of reliability based on the variance of localization. This measure is used to determine the relative weightings of vision and audition in both integration and adaptation. These models therefore suggest that if the reliability of the modalities were experimentally manipulated, the weights and effective learning rates would adapt. In this experiment we explored one aspect of this hypothesis by introducing variance into the visuo-auditory relationship and assessing any changes in localization.

Experiments on adaptation to disarrangement (i.e. varying perturbations) have a long history in the visuomotor system, although studies in the auditory modality are few. Adaptation to disarrangement was first studied by Cohen & Held (1960) who showed that exposure to a prism, cyclically varying in displacement from $+22°$ to $-22°$ at a rate of 1 cycle every 2 min, failed to elicit visuomotor adaptation. However, if the subject produced active limb movements during prism exposure, the variability in pointing increased. Similar results were found for random (non-cyclical) visuomotor perturbations (Efstathiou, 1963; Abplanalp and Held, 1965).

Freedman and colleagues conducted experiments on auditory disarrangement in which sounds with random interaural time differences were paired with head movements (e.g. Freedman & Pfaff, 1962; Freedman & Zacks, 1964; reviewed in Welch, 1978). This led to an increase in the variability in localizing unseen sounds after exposure during active movements, but no change after exposure in a passive condition. It is hard to evaluate these results, as there is no reason the sensory system could interpret these signals as variable signals from one locus, rather than signals from many fixed loci or one moving locus—it is therefore not clear that the sensory system had any cue for rearrangement (Welch, 1978).

In the following experiment subjects were exposed to a randomly-varying displacement between the visual and auditory locations of visuo-auditory stimuli. The perturbation consisted of zero mean, constant variance noise added to the location of one stimulus (e.g. the auditory) in relation to the other stimulus (e.g. the visual).[10]

2.7.1 Method

Subjects

Eight right-handed subjects (5 male, 3 female; ages 18-27) participated in this experiment. Subjects were naive to the purpose of the experiment, gave their informed

[10] Nothing in the experiment distinguished whether the noise was added to vision or audition—noise was added to the relation *between* the two.

consent, and were paid $7.00 for participation. All subjects had self-reported normal or corrected-to-normal vision and normal hearing.

Paradigm

Except for the nature of the perturbation, the paradigm was identical to the one used in Experiment 2. For all subjects the perturbation was zero-mean, 10° standard deviation (s.d.) Gaussian noise added to the relation between vision and audition. That is, on each trial a random number was independently generated from the Gaussian distribution, and used as a displacement. The range of the visual stimuli was kept constant while the range of auditory stimuli accommodated the perturbation. Perturbations were cut off at 2 s.d. (20°) to avoid very large displacements that would fall outside the experimental range or make the subject conscious of the discrepancies. The actual variance of the truncated Gaussian noise was therefore reduced to 77.5 deg^2.

As in Experiment 2, the perturbation was absent for the first 3 sessions (trials 1–105; pre-exposure), was increased linearly during the 4^{th} session (trials 106–140), and was present in-full for the last 8 sessions (trials 141–420). Only a third of the trials were visuo-auditory; the purely visual and auditory trials throughout the experiment could therefore be used to assess adaptation.

2.7.2 Results

Adding 10° s.d. noise to the relationship between visual and auditory stimuli did not significantly change the overall variance of visual or auditory localization (Figure 2-10a & b). Specifically, comparing sessions 6-12 to sessions 1-3, visual variance increased by 0.66 ± 0.70 (ns), and auditory variance increased by 0.56 ± 5.6 (ns). On the other hand, the variance of localizing visuo-auditory stimuli, which was computed relative to the location of the visual stimulus, increased significantly by 2.23 ± 0.68 ($p < 0.01$). This corresponds to an increase of 60 % over baseline (Figure 2-10c).

To measure the reliability of baseline variances computed from sessions 1-3, they were compared to the average variances in Experiment 1 for the same group of subjects. For visual, auditory, and visuo-auditory stimuli, the baseline variances were not significantly different from the variances in Experiment 1—the differences were 0.21 ± 0.59, 4.7 ± 5.4, and 0.74 ± 0.63, respectively for the three types of stimuli.

The spatial pattern of localization variance for the three conditions, though hard to interpret due to the inherent measurement noise, showed an interesting pattern on the left side. While the variance of visual and visuo-auditory localization increased, the variance of auditory localization decreased (Figure 2-11a, b & c; observational results, no significance test). Furthermore, the optimal mixing function computed from these variances (c.f. equation (2.23)) decreased significantly relative to the baseline ($p <$

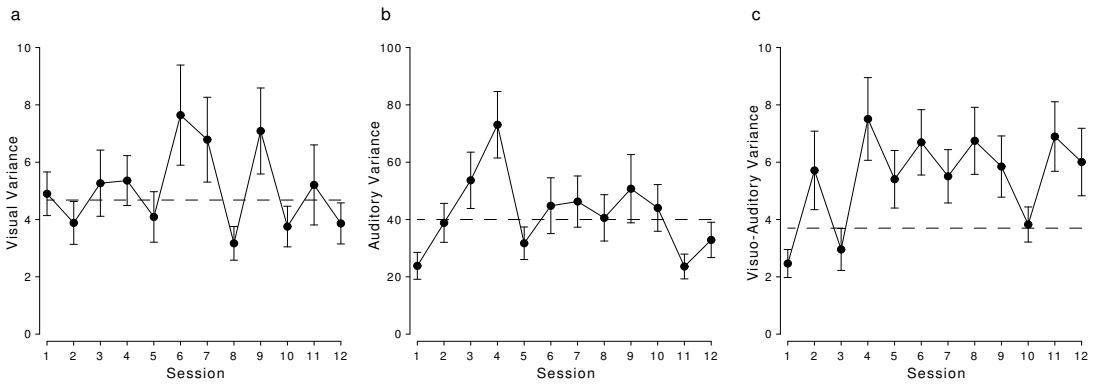

Figure 2-10: Variance as a function of session for a) visual, b) auditory and c) visuo-auditory stimuli. Means (solid circles) and standard errors were computed by averaging the variance over each session; the baseline variances (dashed lines) were computed by averaging over sessions 1-3.

0.01),[11] indicating an overall change in the relative proportions of visual and auditory variance (Figure 2-11d).

2.7.3 Discussion

Increasing the variance in the relationship between vision and audition did not increase the overall localization variability in either modality.[12] The overall variability in pointing to visuo-auditory stimuli did, however, increase. Although this increase was substantial (2.23 ± 0.68; 60% of baseline), it was minuscule compared to the perturbation (2.9% of the 77.5 deg^2 added variance). Since the perturbation was in effect during visuo-auditory localization, an increase in variance indicates that despite strong visual capture the auditory stimulus had an effect on localization.

The spatial pattern of changes in variance seems to indicate that although the variance of visual and visuo-auditory localization increased slightly in some regions, the variance of auditory localization decreased in those same regions. This effect is translated into a significantly lower estimated optimal weighting for vision. Such a change in weighting would have two consequences: (1) a decrease in the magnitude of visual capture, and (2) an increase in visual adaptation. The results in this experiment do not address these predictions. However, both these predictions can be tested

[11] At each location, the differences and standard errors were used to compute a Z score. The mean Z score was 0.46 ± 0.16, significantly different from zero.

[12] Contrary to the findings of Freedman and colleagues (Freedman and Pfaff, 1962; Freedman and Zacks, 1964), we did not observe an increase in the variability of pointing to auditory stimuli. However, as we have already mentioned, the methodology and assumptions of Freedman's studies make direct comparison difficult.

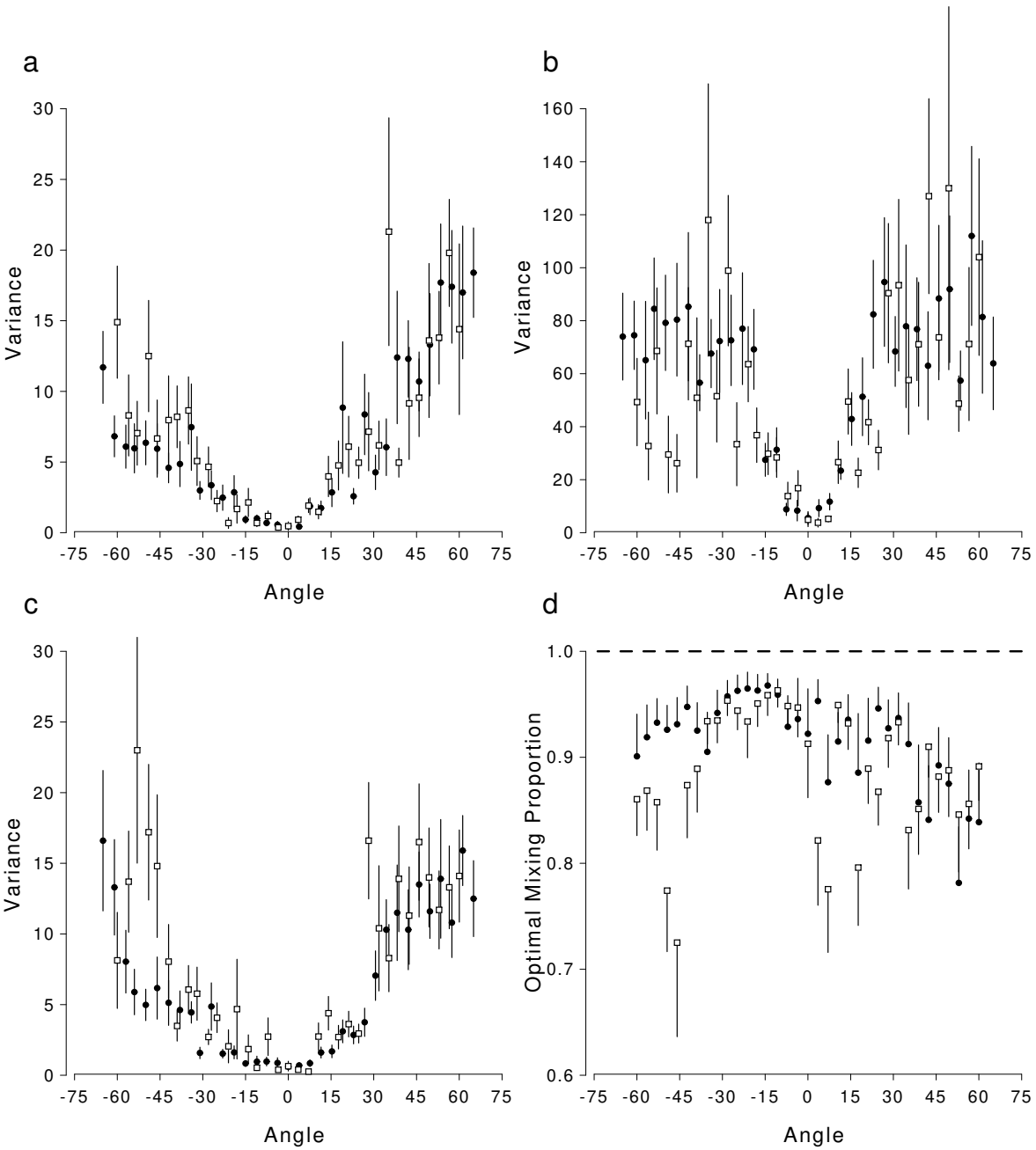

Figure 2-11: Variance of localization plotted as a function of stimulus location (mean $\pm$ 1 s.e.) for a) visual, b) auditory, and c) visuo-auditory stimuli. The baseline condition (solid circles) computed from Experiment 1 is shown along with the post-exposure condition (white squares). d) Optimal mixing proportion computed from baseline (solid circles, upward error bars) and post-exposure (white squares, downward error bars) variances.

in experiments where adaptation to added variance is followed by exposure to a consistent visuo-auditory bias.

Model predictions

The minimum variance model predicts that the variance in pointing to visuo-auditory stimuli will be

$$w_{\text{vis}}^2\sigma_{\text{vis}}^2 + w_{\text{aud}}^2\sigma_{\text{aud}}^2.$$

In the analysis, we computed visuo-auditory variance relative to the location of the visual stimulus; for this measure, the experimentally added variability is included in the variance of the auditory stimulus. The increase in visuo-auditory variance predicted by the minimum variance model is therefore given by $w_{\text{aud}}^2(\Delta\sigma_{\text{aud}}^2)$. Assuming a range of visual weighting from 0.8 to 0.95, this gives a predicted increase in variance ranging from 0.2 to 3.1 deg^2, comparable to the 2.23 deg^2 observed. In contrast, the competitive model predicts no change in variance, and the stochastic model predicts, for the same range of visual weightings, a 3.9 to 15.5 deg^2 increase.

So far, we have discussed predictions regarding integration—i.e. the immediate changes in pointing resulting from visuo-auditory variability. To understand any more permanent adaptive effects of increasing intersensory variability within the framework of optimal estimation, we recall that the weighting given each modality is inversely proportional to its estimated localization variance. Increasing variability should increase these estimated variances and may therefore alter the weighting of the modalities. Which modality the increased variability is attributed to determines which direction the weighting will change. For example, if the added variability is attributed mostly to audition, visual dominance will increase. However, just as it is impossible to know which sense is "correct," it is also impossible to know which to attribute the variability to. Nevertheless, the optimal estimation framework suggests a means of estimating the amount of increased variance attributable to each sense.

The combined variability between the senses, $\sigma_{\text{vis-aud}}^2$, is assumed to originate from the variability in each sense, i.e. under the independent noise assumption:

$$\sigma_{\text{vis-aud}}^2 = \sigma_{\text{vis}}^2 + \sigma_{\text{aud}}^2.$$

We wish to obtain estimates of σ_{vis}^2 and σ_{aud}^2, denoted $\hat{\sigma}_{\text{vis}}^2$ and $\hat{\sigma}_{\text{aud}}^2$, from this combined intersensory variance. Noting that the MVE weights, w_{vis} and w_{aud} are defined as the proportions of the intersensory variance attributed to audition and vision respectively, it is clear that the only self-consistent estimates of the variances are:

$$\begin{aligned}\hat{\sigma}_{\text{vis}}^2 &= w_{\text{aud}}\sigma_{\text{vis-aud}}^2 \\ \hat{\sigma}_{\text{aud}}^2 &= w_{\text{vis}}\sigma_{\text{vis-aud}}^2.\end{aligned}$$

Any other choice for the proportion attributed to each variance would result in a contradiction when $\sigma^2_{\text{vis-aud}} = \sigma^2_{\text{vis}} + \sigma^2_{\text{aud}}$, i.e. when there is no experimentally induced variability. If the intersensory variability is increased experimentally by σ^2_{exp}, such that $\sigma^2_{\text{vis-aud}} = \sigma^2_{\text{vis}} + \sigma^2_{\text{aud}} + \sigma^2_{\text{exp}}$, then each estimated variance will be increased,

$$\begin{aligned} \Delta\hat{\sigma}^2_{\text{vis}} &= w_{\text{aud}}\sigma^2_{\text{exp}} \\ \Delta\hat{\sigma}^2_{\text{aud}} &= w_{\text{vis}}\sigma^2_{\text{exp}}. \end{aligned}$$

The added variance will be mostly attributed to the already less reliable sense, and least to the most reliable sense. The weighting of the modalities will, however remain unchanged,

$$w^{\text{new}}_{\text{vis}} = \frac{\hat{\sigma}^2_{\text{aud}} + \Delta\hat{\sigma}^2_{\text{aud}}}{\hat{\sigma}^2_{\text{aud}} + \hat{\sigma}^2_{\text{vis}} + \sigma^2_{\text{exp}}} = \frac{w_{\text{vis}}(\hat{\sigma}^2_{\text{aud}} + \hat{\sigma}^2_{\text{vis}}) + w_{\text{vis}}\sigma^2_{\text{exp}}}{\hat{\sigma}^2_{\text{aud}} + \hat{\sigma}^2_{\text{vis}} + \sigma^2_{\text{exp}}} = w_{\text{vis}}.$$

This model therefore predicts that despite an increase in the estimated variances of the two modalities the weighting will not change. Again, the results in this experiment do not address this prediction. This prediction can be tested in an experiment where subjects are presented with a visuo-auditory displacement after adaptation to added variance.

Summarizing, the changes in actual visual and auditory localization variance suggest that the optimal weighting between the senses should decrease. A model based on minimum variance integration, however, suggests that while the estimates of variance in each modality should increase, the weighting should remain invariant. Testing between these two alternatives requires further experiments. Finally, a large (60%) increase in visuo-auditory variance was observed experimentally. This increase is consistent with the minimum variance model of integration, but falls outside the ranges predicted by the competitive and stochastic models.

2.8 Experiment 4: Generalization of the Visuo-Auditory Map

In this experiment we examine adaptation at loci other than the locus of exposure, a form of adaptation known as *generalization*. The paradigm limits concurrent visuo-auditory exposure to a single point. By displacing the relation between vision and audition at that point and testing visual and auditory localization at other points, the pattern of generalization can be assessed.

The generalization paradigm addresses two sets of issues. First, like Experiment 2

the results can be used to distinguish between different computational models of integration and adaptation. The different models predict varying extent of adaptation depending on the locus of exposure and testing, the weightings of the modalities, and other factors. Second, the generalization paradigm can be used to infer properties of the representations underlying visual and auditory maps of space. This topic is discussed at length in Chapter 4, which is dedicated exclusively to generalization patterns in the visuomotor coordinate transformation.

2.8.1 Method

Subjects

Eight right-handed subjects (5 male, 3 female; ages 18-27) participated in this experiment. Subjects were naive to the purpose of the experiment, gave their informed consent, and were paid $7.00 for participation. All subjects had self-reported normal or corrected-to-normal vision and normal hearing.

Paradigm

Except for the nature of the perturbation, the paradigm was identical to the one used in Experiment 2. As before, one third of the trials were visuo-auditory. During these trials the visual stimulus was always located at 24.7° (the 9^{th} target from the right). The concurrent auditory stimulus started at this location for the first 3 sessions, linearly shifted 15° to the left to 9.7° during the 4^{th} session, and remained at 9.7° for the rest of the experiment. Thus concurrent visual flashes and auditory buzzes were limited to a single visual location 24.7° .

2.8.2 Results

Significant changes in auditory localization occurred after the one-point visuo-auditory remapping (Figure 2-12a). Generalization was most pronounced in the right periphery (15° to 45°) reaching up to 9.5° (63% of the displacement) at 38° in the azimuth, declined to zero near the locus of remapping (9.7°) and straight ahead (0°), and continued declining below zero for about 30° beyond this point (Figure 2-12b). The pattern in the left periphery (–60° to –30°) was more erratic.

The time course of spatial generalization was analyzed by comparing changes in auditory localization relative to baseline during three different phases of exposure: sessions 4-6, 7-9, and 10-12 (Figure 2-13). Generalization gradually increased and spread from right (45°) to left (-30°), though it was never significantly different from zero at the remapped point or at 0° .

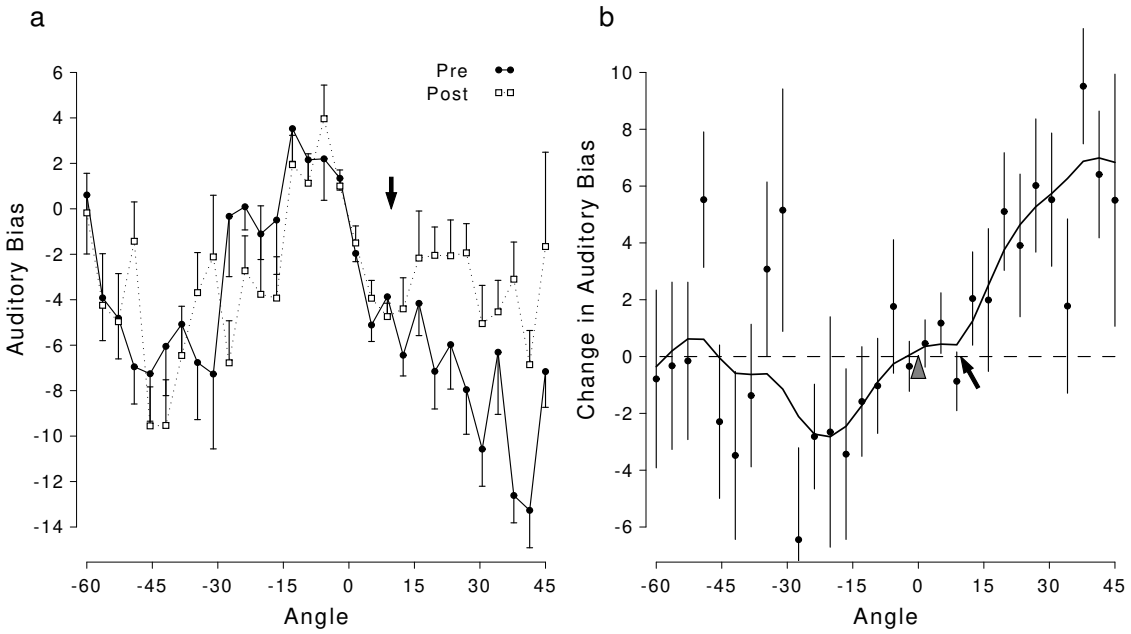

Figure 2-12: Pattern of generalization. a) Auditory bias for baseline (black circles; calculated from trials 176–420 of Experiment 1), and post-exposure auditory bias (white squares; calculated from trials 176–420 of Experiment 4). b) Change in auditory bias computed from a) and fitted with an 8 degree of freedom smoothing spline. The positive direction indicates adaptive changes. The black arrows indicate the locus of the exposure. The grey triangle in b) marks 0° .

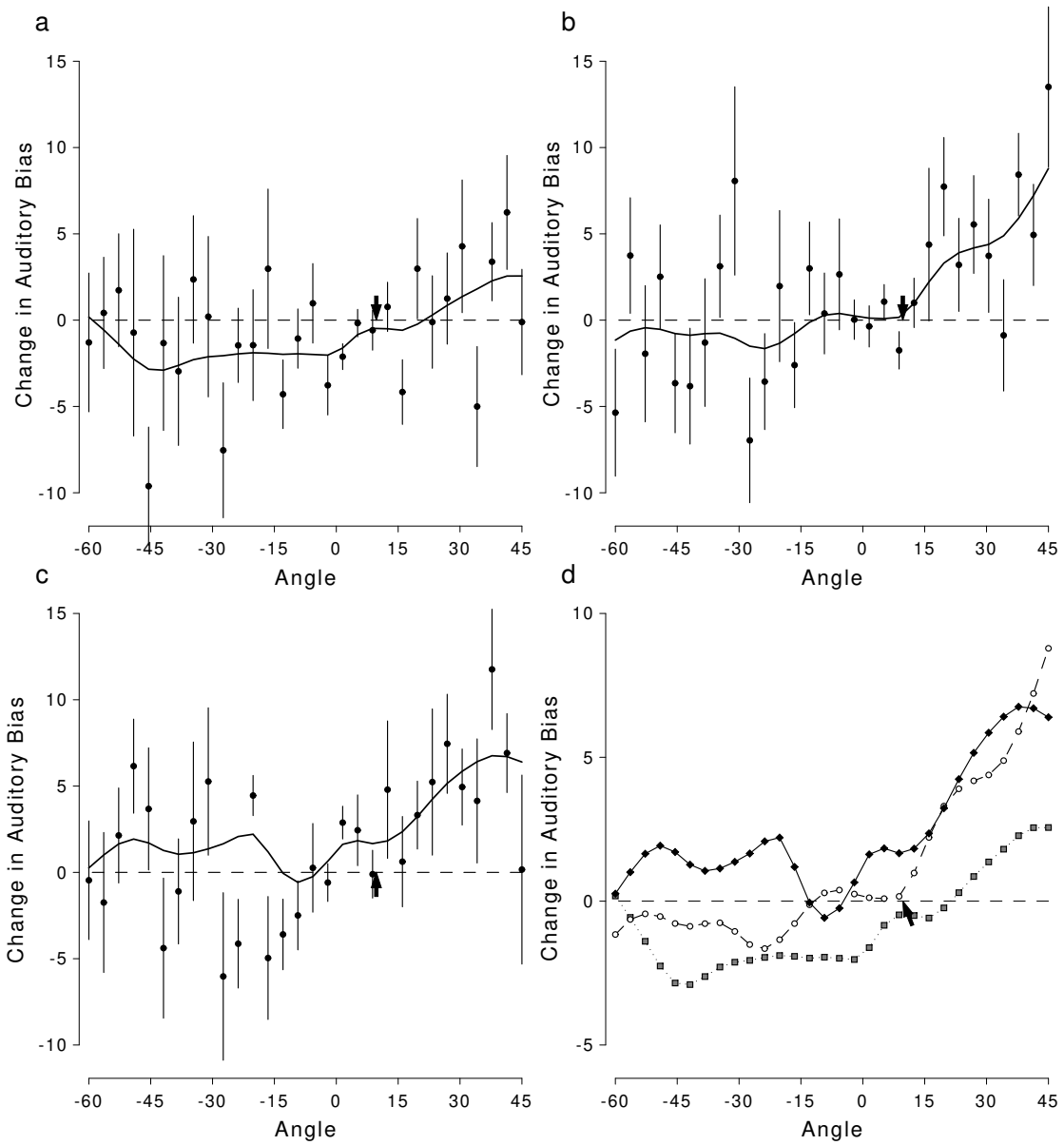

Figure 2-13: Time course of generalization, computed by subtracting the baseline bias from the the bias for a) sessions 4-6 (trials 106–210), b) sessions 7-9 (trials 211–315), and c) sessions 10-12 (trials 316–420), with 8 degree of freedom smoothing spline fits. d) Superimposed smoothed fits from a) (grey squares), b) (white circles) and c) (black diamonds). The positive direction indicates adaptive changes, and the black arrow indicates the locus of exposure.

2.8.3 Discussion

Exposure to a local 15° displacement caused no shift in auditory localization at the remapped point. Significant and increasing shifts did, however, arise in the periphery, on the side of the remapping. Furthermore, although no shift occurred straight-ahead, localization shifted in the direction *opposite* to adaptation on the side opposite the remapped point.

Taken together, these results suggest the following account: (1) Adaptation is least around 0° and larger towards the periphery. (2) The constraint to adapt least around 0° precluded adaptation at the nearby remapped point (9.7°), however, (3) the effects of the displacement did generalize to the more adaptive right periphery. Finally, (4) the combined effect of the shift on the right and the lack of shift straight-ahead resulted in an expansion of the auditory map on the right side.[13] This expansion generalized to the left side, resulting in the negative shifts observed (Figure 2-12b).

We address each of these points in turn. Points (1) & (3) are consistent with the finding in Experiment 2 that adaptation to a constant displacement was smallest around 0°and larger in the periphery. The results from this generalization experiment, therefore, also support the variance-weighted delta rule (VWDR) model of adaptation, which, in contrast to the DR and the WDR models, predicts much smaller adaptation straight-ahead. Point (2) reflects an assumption of smoothness in the visuo-auditory relationship. We assume that unless exposed to a severe perturbation,[14] the mapping between points in visual and auditory space is represented smoothly. Therefore, if the map is constrained to shift very little at 0° , it is unlikely to shift much at 9.7° .

To understand how a shift at one location could result in an expansion of the auditory map (point 4), it is important to note that a perturbation at one point is consistent with many possible visuo-auditory remappings. To accommodate a one-point perturbation, the CNS could remap the visuo-auditory relationship through:

1. a local shift of the mapping at the perturbed point, preserving the natural mapping elsewhere,

2. a semi-local shift at the perturbed point and points nearby in space,

3. a semi-local shift at the perturbed point and points nearby in some acoustic representation other than space (e.g. same ITD),

4. a global displacement of the visuo-auditory relation along the whole azimuth,

5. a global expansion/contraction of the visuo-auditory map, etc.

[13] We call this an *expansion* because localization right of 0° shifted in the rightward (positive) direction.

[14] An example of such a perturbation is the complete elimination of visual input from birth in barn owls (Knudsen et al., 1991). This perturbation caused tectal auditory maps to develop erratically.

Generalization studies are based on the hypothesis that the pattern of adaptation that emerges reflects intrinsic constraints of the representation (Bedford, 1989; Shadmehr & Mussa-Ivaldi, 1994; Chapter 4). The pattern that we found in this study, visuo-auditory shifts growing from center to periphery, suggests that the more accurate central region is constrained to be less adaptable. The expansion found may also be a consequence of the underlying representation of auditory space. An expansion or contraction around straight-ahead could be represented as a scaling of interaural time or intensity difference cues. This simple mechanism could partially account for the adaptation found.[15]

2.9 Controls

2.9.1 Alternative Cues to Auditory Stimulus Location

One of the concerns in the experimental design was that the subjects may have made use of cues other than the actual auditory stimulus for localization. For example, if the subject were able to see through the white screen or discern the location of the rod which the speaker was mounted on using the noise from the stepper motor, the auditory localization task would be corrupted by these extraneous cues. To test for this possibility we conducted a control experiment in which subjects were instructed to use any cues available to them to guess, as best as possible, the location of the end of the rod—neither an auditory nor visual stimulus was presented.

Four subjects (2 male, 2 female; ages 21-29) participated in this control experiment. All subjects had already participated in Experiment 1, were explained the apparatus and the purpose of the control experiment. The experimental procedures were the same as in Experiment 1, except: (1) the experiment consisted of only 35 trials (one per stimulus location), (2) neither a visual nor auditory stimulus was presented.

Table 2.2 summarizes the results of the experiment. All subjects reported that the task was very difficult and that they were guessing the target location. The average errors ranged from 50° to 68° . For three out of four subjects there was no correlation between actual rod location and the location pointed to; for one subject there was a negative correlation.

[15] A more speculative interpretation of the expansion of auditory localization responses comes from direct analogy to studies of remapping in the somatosensory system. Recanzone, et al (1992) showed that, upon repeated and attended-to tactile stimulation of the hand, both the receptive field size and the cortical representation of the stimulated region in area 3b increased significantly in owl monkeys. In our experiment, the exposure phase repeatedly stimulated one location of auditory space. It is therefore possible that the representation of this location increased in size, resulting in an expansion relative to the corresponding area of visual space. This would suggest that the expansion was unrelated to the displacement—an easily testable hypothesis.

Subject	Slope	Intercept	R^2	$F_{1,33}$	Prob(F)	RMS Error
AC	0.021	96.0	0.0006	0.021	0.887	49.8°
DW	0.097	81.8	0.0091	0.303	0.586	53.7°
WY	-0.165	116.1	0.0405	1.392	0.247	56.2°
WZ	-0.649	144.3	0.3801	20.235	0.0001	68.0°

Table 2.2: Summary of alternative cues control experiment.

These results suggest that motor noises, sight of the rod through the screen, or other extraneous cues could not contribute significantly to localization accuracy in Experiments 1 to 4.

2.9.2 Pointing with the Left Hand

Localization of visual and auditory stimuli was measured in all the experiments using a pointing paradigm. The bias and variance obtained through this procedure is surely a contribution both of errors in localization and variability in the pointing response—i.e. sensory and motor errors. There are several ways in which the variability in pointing can be factored out, obtaining a more accurate measure of localization. First, an alternative measure of localization can be used, for example through as eye movements, and the results compared to pointing responses. Based on some statistical assumptions, such as additivity of sensory and motor noise, the purely sensory component of the bias and variance of localization can be estimated. Alternately, a purely sensory paradigm for localization could be used. For example, in a two alternative forced choice (2AFC) paradigm subjects would be presented with target and probe stimuli and asked to judge whether the target is left or right of the probe. This paradigm, however, would require far too many presentations of stimuli to estimate localization over the azimuth.

In this control we simply sought to estimate the effect on localization due to pointing with the right hand. Several of the effects observed in Experiments 1 to 4 were asymmetrical about 0° . How much of this asymmetry was due to the pointing response?

Five right-handed subjects (2 male, 3 female; ages 19-27) participated in this control experiment. All subjects had already participated in Experiment 1, and were familiar with the experiment. This experiment was identical to Experiment 1 except that subjects pointed to the stimuli using a marker worn on the *left* index finger.

Figure 2-14 compares localization bias when pointing with the left and right hands for the same set of subjects. For visual, auditory, and visuo-auditory stimuli, bias is shifted uniformly in the rightward (positive) direction for left-handed pointing. This indicates a relative overshoot from the starting point of movement. (Note that left-handed pointing movements started from the left side, and right-handed movements started from the right side). The asymmetries in bias were approximately reversed

with handedness, suggesting that handedness could account for the asymmetric biases found in Experiments 1 to 4. The effect of handedness was large (up to 15°) in the periphery and generally vanished straight-ahead. The pattern of effects was also different for pointing to visual and auditory stimuli (i.e. there was an interaction of modality and handedness).

Figure 2-15 compares the variance of localization when pointing with the left and right hands for the same set of subjects. The pattern of variance—smallest straight-ahead and increasing to the periphery—is present for all three stimulus modalities for both left and right handed pointing. The variance of right handed pointing seems generally smaller than the variance of left handed pointing, especially in the periphery.

Summarizing, handedness of pointing has an effect on both the bias and variance measures of localization. Although the effects on bias are significant, the effects on variance are relatively small and preserve the pattern of smallest straight-ahead, largest in the periphery. This suggests that the predictions of the models in this chapter, which are all based on measures of relative localization variance, remain effectively unaffected by which hand the subject pointed with.

2.10 Discussion

In this chapter we first posed the problems of intersensory integration and intersensory adaptation within a computational framework based on statistical estimation. Within this framework the two problems are closely tied—the pattern of adaptation can be predicted from the principle used to integrate two discrepant sensory signals. Three explicit computational models of the integration and adaptation of visual and auditory spatial maps were then proposed and tested through a series of experiments.

2.10.1 Empirical findings

The most striking feature of localization errors for visuo-auditory stimuli is that they are virtually identical to localization errors for visual stimuli (Experiment 1). Furthermore, subjects were generally unaware when large (15°) discrepancies were imposed between vision and audition, a finding which is consistent with the often reported phenomenon of visual capture or "ventriloquism." The variance of visuo-auditory localization was slightly smaller than the variance of visual localization, a finding which is inconsistent with the stochastic model of integration. The extent of visual capture, however, did not allow exclusion of wither the minimum variance and competitive models.

When a displacement was imposed between the two senses, the visual modality did not perceptibly adapt (Experiment 2). However auditory localization adapted significantly, shifting by about 40% of the perturbation. Adaptation was least straight-ahead and larger in the periphery. This finding is consistent with the variance-

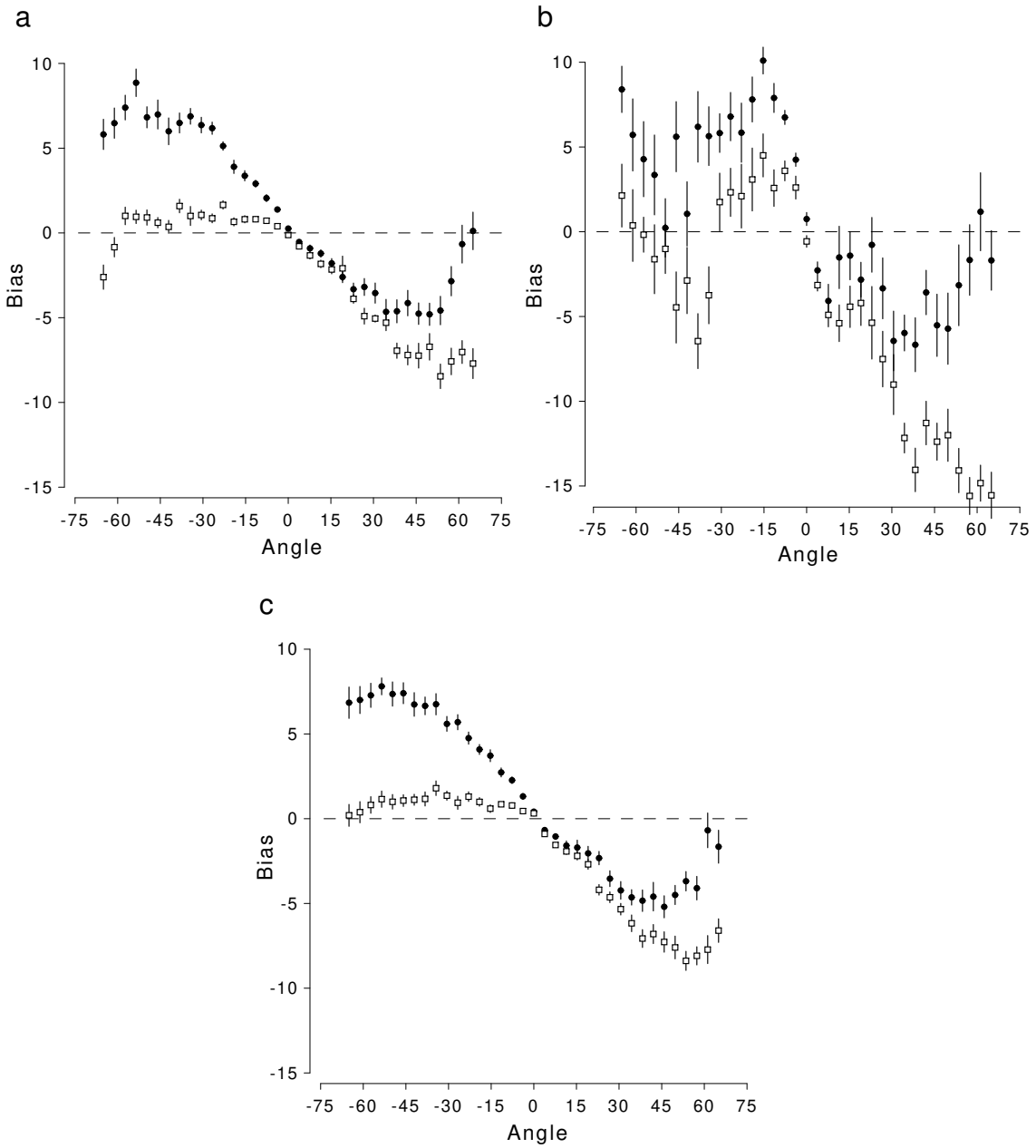

Figure 2-14: Bias of localization plotted as a function of stimulus location (mean ± 1 s.e.) for a) visual, b) auditory, and c) visuo-auditory stimuli. The left-handed pointing control (solid circles) is shown along with the right-handed pointing baseline from Experiment 1 (hollow squares).

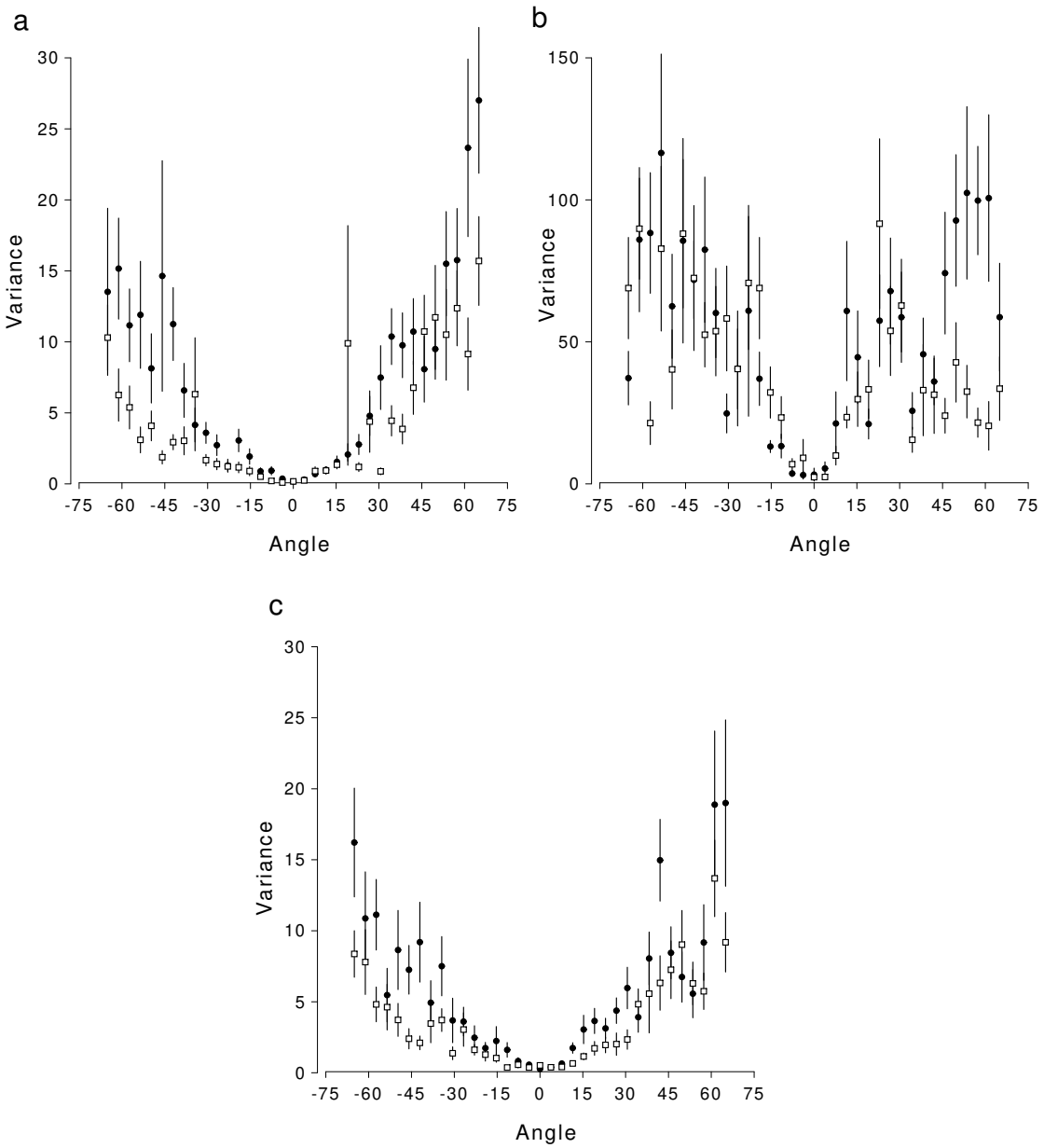

Figure 2-15: Variance of localization plotted as a function of stimulus location (mean $\pm$ 1 s.e.) for a) visual, b) auditory, and c) visuo-auditory stimuli. The left-handed pointing control (solid circles) is shown along with the right-handed pointing baseline (hollow squares).

weighted delta rule (VWDR) for adaptation, and inconsistent with the other two proposed models. The VWDR can be derived from the minimum variance integration model.

Adding substantial variability to the relation between vision and audition had little overall effect on visual and auditory localization, but significantly increased the variance of visuo-auditory localization (Experiment 3). The increase in visuo-auditory variance was consistent with the minimum variance model of integration, but outside the range predicted by the stochastic and competitive integration models.

The pattern of generalization to a local remapping was unexpected in that a 15° displacement induced virtually no adaptation at the locus of exposure (Experiment 4). However, points up to 40° away showed significant adaptation consistent with an expansion of auditory localization about straight-ahead. This finding is again consistent with the VWDR model, as an expansion is a pattern which can both account for some of the perturbation and maintain straight-ahead localization relatively unadapted as VWDR predicts. Furthermore, this pattern suggests that contractions and expansions of auditory space may be simply represented. One of the mechanisms that could subserve this form of auditory adaptation is a simple scaling of the ITD or IID cues.

2.10.2 Implications

These findings suggest that signals from multiple sensory modalities are integrated in such a way that the combined signal has minimal variance. A closely related learning rule—the variance-weighted delta rule—acts to resolve long-term intersensory discrepancies. The variance-weighted delta rule states that the rate of adaptation is proportional to the variability (or inversely proportional to reliability) of each sense. This has several implications: (1) In the limit of complete adaptation to a discrepancy, both senses will converge at a point which is the optimal (minimum variance) fusion of the two.[16] (2) Unlike the weighted delta rule, the VWDR is a local criterion. In other words, the learning rate for each modality does not depend on the variance of the other modality. (3) Some signal must code the reliability of a modality and thus gate learning. The neural code for the reliability of a signal could be explicit. For example, the firing rate of a neuron in a spatial map could be proportional to that neuron's "confidence" that there is a stimulus in that location. On the other hand, the reliability could be coded implicitly. For example, the size of receptive fields is a parameter that could be related both to the variance in localization and to the rate of plasticity (e.g. larger receptive fields are more common in the periphery and suggest a greater pattern of connectivity). These issues cannot be answered at the psychophysical level and must consequently rely on neurophysiological studies.

[16] Of course this limit of complete adaptation is rarely observed in experiments.

2.10.3 Directions for future work

To study integration it is advantageous to have a system in which no single input dominates. Although the visuo-auditory localization system is dominated by vision, it is possible to extend the paradigms used here to create a more balanced compromise between the senses. One way in which this can be done is by manipulating the properties of the stimuli, such as luminance or sound frequency spectrum, or their relative timing. Under appropriate conditions, a more balanced mixing function should arise, setting the stage for tests of both clear visuo-auditory integration (i.e. localizing a visuo-auditory stimulus well between the visual and auditory stimulus) and visual adaptation.

It would be interesting to repeat the experiments in this chapter with the added manipulation of having subjects look at a fixation point to one side, keeping their head straight ahead. This may resolve whether the reduction of auditory localization variance straight ahead is due to auditory cues or to the modulatory effects of eye position and attention. The adaptation experiments may also benefit from this manipulation. For example, by manipulating both the natural displacement of eye-centered and head-centered coordinates that occurs with off-center fixation and the experimentally imposed perturbation, the effective visuo-auditory displacement can be carefully controlled. Different forms of adaptation may occur if the experimental displacement cancels or magnifies the eye-position dependent displacement.

Another promising paradigm for studying localization would replace the pointing response with eye movements. This paradigm has two advantages: First, data from many more localization trials can be collected since eye movements are fast and virtually effortless compared to arm movements. Second, eye movements are more clearly mediated by the superior colliculus than arm movements. The analogous experiments with eye movements could therefore be directly related to the body of literature on the superior colliculus.

2.10.4 Conclusion

When the normal relation between visual and auditory space is altered through an experimentally-induced remapping, the pattern of auditory adaptation that emerges can be predicted though a simple learning rule. This rule states that each sensory modality and each location in space adapts in inverse proportion to its localization acuity. This learning rule is closely tied to the principle of minimum variance integration, which states that the inputs from multiple modalities are integrated so as to maximally reduce uncertainty in the sensory estimate. The pattern of pointing to concurrent visual and auditory stimuli also supports the minimum variance integration principle, providing converging evidence. The problems of adaptation to an intersensory discrepancy, and integration of multisensory inputs are therefore closely tied.

Chapter 3

An Internal Model for Sensorimotor Integration

3.1 Introduction

The ability to reach for a cup or balance on one foot requires the integration of information from several sensory and motor sources. One of the key roles that this sensory information plays is to provide an estimate of the system's state.[1] For example, reaching for a cup requires knowledge of the initial position and orientation of the hand, and balancing requires knowledge of the precise orientation of the body. Lack of knowledge of the initial state of the limb, for example as a result of sensory neuropathy, can result in large movement errors (Ghez et al., 1990, 1995; Gordon et al., 1995).

In this chapter, we study the process of sensorimotor integration involved in estimating the state of a limb during movement. During reaching, the position of the hand can be derived from visual inputs, proprioceptive inputs, and the motor commands issued by the CNS. We examine the propagation of errors in estimating the hand's state as a function of movement duration and externally imposed forces. This error propagation is compared with the predictions of a model based on minimum variance integration (Chapter 2).

The model, an optimal linear observer known as the Kalman filter, estimates the state of the system by monitoring its inputs (the motor commands) and its observable

[1]The *state* is defined as the set of variables which, when known, make predicting the future behavior of a system independent of knowledge of the past behavior. For a mechanical system, for example, the state is generally defined as the positions and velocities of all its components. Given the current state, the future states are independent of the past states.

outputs (the visual and proprioceptive signals arising from the movement). The current estimate of the state is derived by simulating the forward dynamics of the system using the previous estimate of the state and the perceived motor command (c.f. equation 2.14). The component of the observer which simulates the dynamics of the controlled process is known as an internal model.

Based on computational principles alone, it has been previously proposed that the central nervous system uses an internal model to simulate the dynamic behavior of the motor system in planning, control and learning (Sutton and Barto, 1981; Ito, 1984; Kawato et al., 1987; Jordan and Rumelhart, 1992; Miall et al., 1993). The experimental results and simulations in this chapter provide direct evidence for the existence and use of such an internal model.

3.2 Experiment: Propagation of Errors in Sensorimotor Integration

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 (Jordan, 1995). 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. Forward models—the focus of this article—have been been shown to be of potential use for solving four fundamental problems in computational motor control. First, the delays in most sensorimotor loops are large making feedback control infeasible for rapid movements. By using a forward model for internal feedback the outcome of an action can be estimated and used before sensory feedback is available (Ito, 1984; Miall et al., 1993). Second, a forward model is a key ingredient in a system that uses motor outflow ("efference copy") to anticipate and cancel the reafferent sensory effects of self-movement (Gallistel, 1980; Robinson et al., 1986). Third, a forward model can be used to transform errors between the desired and actual sensory outcome of a movement into the corresponding errors in the motor command, thereby providing appropriate signals for motor learning (Jordan and Rumelhart, 1992). Similarly by predicting the sensory outcome of the action, without actually performing it, a forward model can be used in mental practice to learn to select between possible actions (Sutton and Barto, 1981). Finally, a forward model can be used for state estimation in which the model's prediction of the next state is combined with a reafferent sensory correction (Goodwin and Sin, 1984). Although shown to be of theoretical importance, the existence and use of an internal forward model in the CNS is still a major topic of debate.

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 via the use of a forward model. To test between these possibilities, we carried out an experiment in which subjects made arm movements in the dark. Three experimental conditions were studied, involving the use of null, assistive and resistive force fields. 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 (see Appendix A of this chapter). The bias of this location estimate, plotted as a function of movement duration shows a consistent overestimation of the distance moved (Figure 3-1). This bias shows 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 shows 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 can be fully accounted for only if we assume that the motor control system integrates the efferent outflow and the reafferent sensory inflow. To establish this conclusion we have developed an explicit model of the sensorimotor integration process which contains as special cases all three of the methods referred to above (see Appendix B of this chapter). The model—a Kalman filter (Kalman and Bucy, 1961)—is a linear dynamical system that produces an estimate of the location of the hand by monitoring both the motor outflow and the feedback as sensed, in the absence of vision, solely by proprioception. Based on these sources of information the model estimates the arm's state, integrating sensory and motor signals to reduce the overall uncertainty in its estimate. The model is a combination of two processes which together contribute to the state estimate. The first process uses the current state estimate and motor command to predict the next state by simulating the movement dynamics with a forward model. The second process uses the difference between actual and predicted reafferent sensory feedback to correct the state estimate resulting from the forward model. The relative contributions of the internal simulation and sensory correction processes to the final estimate are modulated so as to provide optimal state estimates. 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.[2] All other components of the internal model were set to their veridical values.

Simulations of the Kalman filter demonstrate the two distinct phases of bias prop-

[2]This is consistent with the independent data that subjects tend to under-reach in pointing tasks suggesting an overestimation of distance traveled (Soechting and Flanders, 1989).

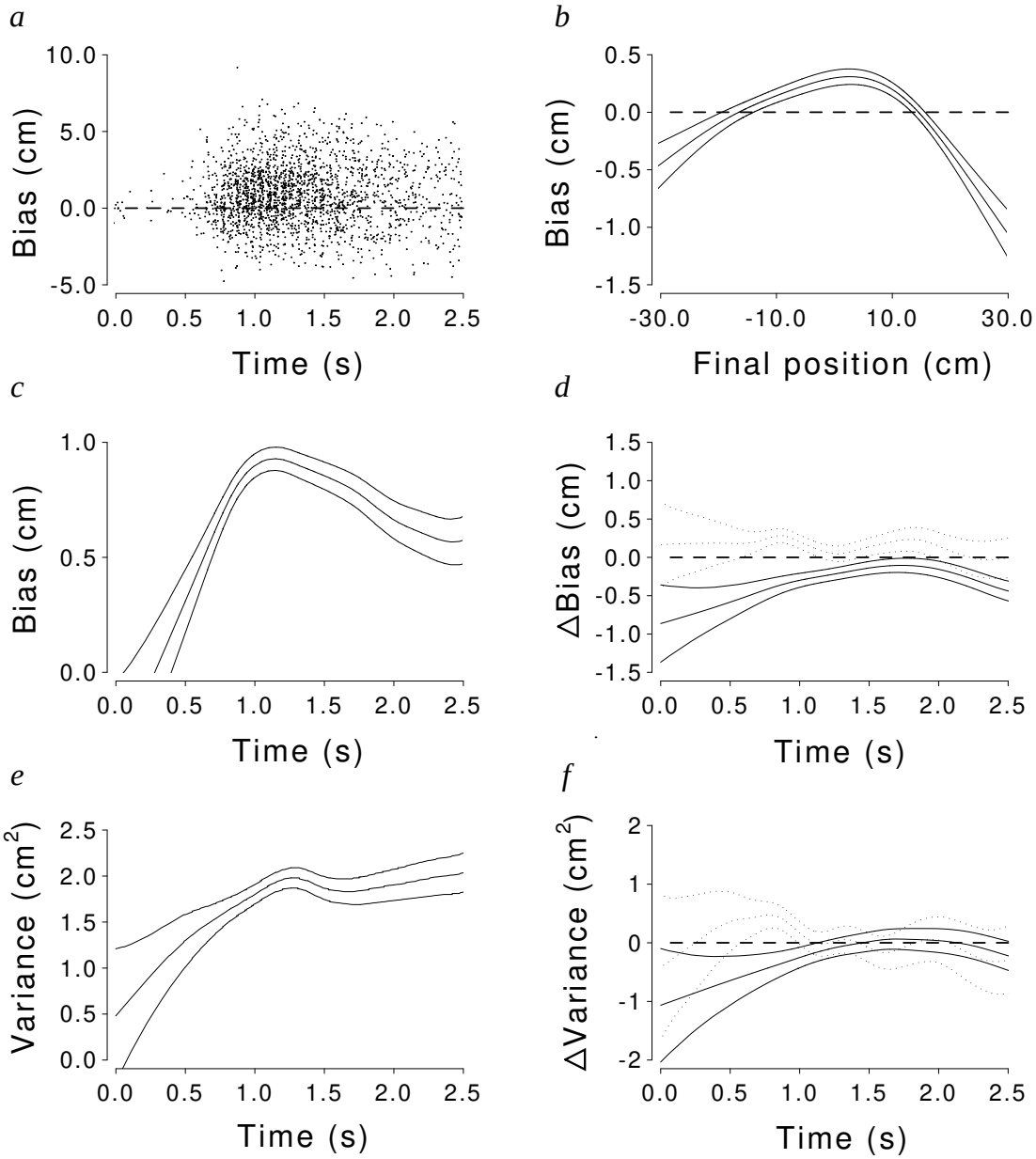

Figure 3-1: (a) Raw data showing localization bias as a function of movement duration. A positive bias represents an overestimation of the distance moved. (b) Mean $\pm$ 1 s.e. fits of bias as a function of final position showing position-dependent kinematic inaccuracies (see Appendix A of this chapter). The propagation of the (c) bias and (e) variance of the state estimate is shown, with standard error lines, against movement duration. The differential effects on (d) bias and (f) variance of the external force, assistive (dotted lines) and resistive (solid lines), are also shown relative to zero (dashed line). 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.

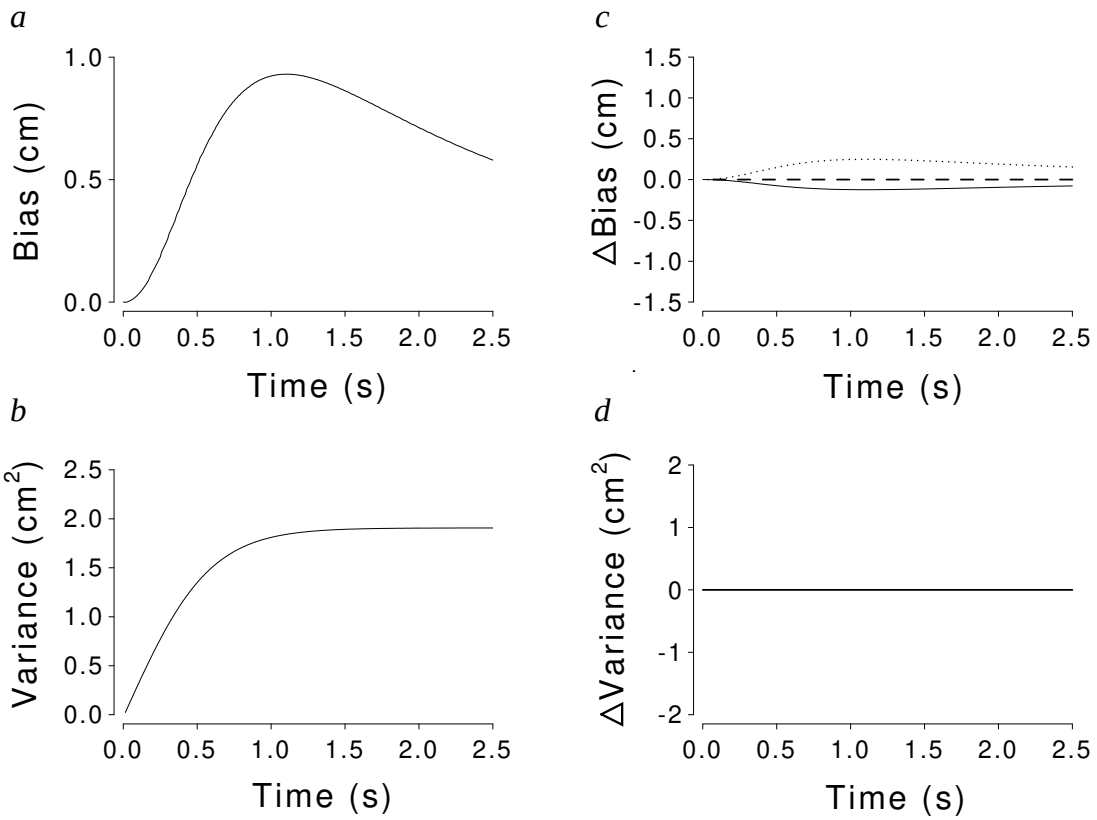

Figure 3-2: Simulated bias and variance propagation from a Kalman filter model of the sensorimotor integration process. (a-d) are in the same representation and scale as (c-f) in the previous figure.

agation observed (Figure 3-2). 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 captures 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. In accord with the experimental results the model predicts no change in variance under the two force conditions.

We have shown 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 Kalman filter 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 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 our 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. We feel that the results of this state estimation study provide strong evidence that a forward model is used by the CNS in maintaining its estimate of the hand location. Furthermore, the state estimation paradigm provides a framework to study the sensorimotor 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 the forward model (Miall et al., 1993).

3.3 Appendix A: Paradigm

The experimental setup consisted of a planar virtual visual feedback system (described in Wolpert, Ghahramani, and Jordan, 1995) in conjunction with a planar two degree-of-freedom torque-motor-driven manipulandum (described in Faye, 1986; see Figure 3-3). The subject gripped a manipulandum on which his thumb was mounted. 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. The virtual visual feedback system was used to project computer-controlled images into the plane of the movement. Eight subjects, who gave their informed consent, participated and performed 300 trials each. 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 perceive visually his initial arm configuration. The light was then extinguished leaving just the initial target. The subject was then required to move his hand either to the left or right, as indicated by an arrow in the initial starting square. This movement was made in the absence of visual feedback of arm configuration. The subject was instructed to move until he heard a tone at which point he stopped. The timing of the tone was controlled to produce a uniform distribution of path lengths from 0–30 cm. During this movement the subject either moved in a randomly selected null or constant assistive or resistive 3N force field generated by the torque motors. Although it is not possible to directly probe a subject's internal representation of the state of his arm, we can examine a function of this state—the estimated visual location of the thumb. The relationship between the state of the arm and the visual coordinates of the hand is known as the kinematic transformation (Craig, 1986). Therefore, once at rest the subject indicated the visual estimate of his unseen thumb position using

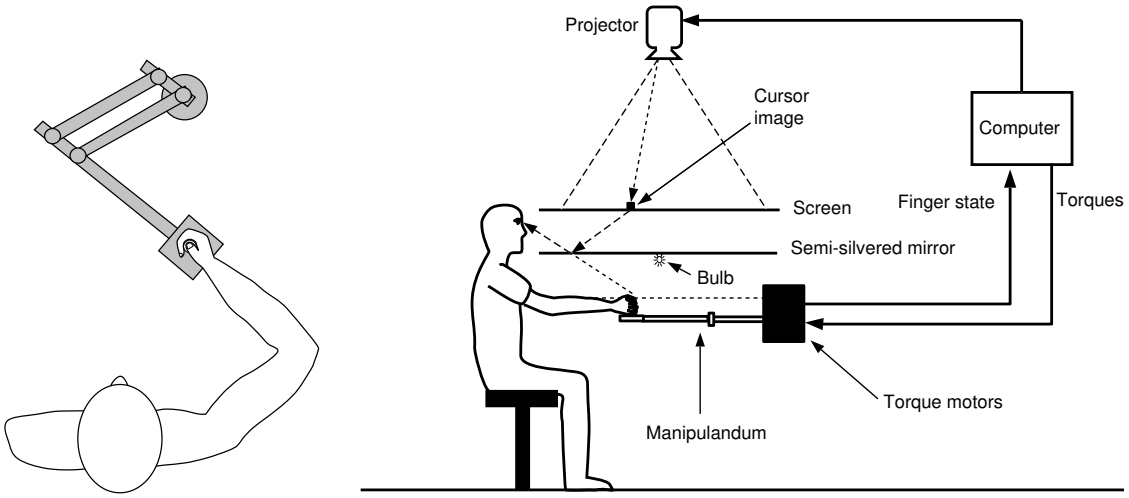

Figure 3-3: Experimental apparatus

a trackball, held in his other hand, to move a cursor projected in the plane of the thumb along the movement line. The discrepancy between the actual and visual estimate of thumb location was recorded as a measure of the state estimation error. The bias and variance propagation of the state estimate was analyzed as a function of movement duration and external forces. A generalized additive model (Hastie and Tibshirani, 1990) with smoothing splines (five effective degrees of freedom) was fit to the bias and variance as a function of final position, movement duration and the interaction of the two forces with movement duration, simultaneously for main effects and for each subject. This procedure factors out the additive effects specific to each subject and, through the final position factor, the position-dependent inaccuracies in the kinematic transformation.

3.4 Appendix B: Simulation

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 $u = u_{\text{int}} + u_{\text{ext}}$, combining both internal motor commands and 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 with an associated covariance matrix given by $Q = E[\mathbf{w}(t)\mathbf{w}(t)^T]$. The system has an observable output, the sensory information, representing the proprioceptive

signals (e.g. from muscle spindles and joint receptors), $\mathbf{y}(t)$ 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 which has the associated covariance matrix $R = E[\mathbf{v}(t)\mathbf{v}(t)^T]$. We assume that this system is fully observable and choose C to be the identity matrix. At time $t = 0$ the subject is given full view of his arm and, therefore, starts with an estimate $\hat{\mathbf{x}}(0) = \mathbf{x}(0)$ with zero bias and variance—we assume that vision calibrates the system. At this time the light is extinguished and the subject must 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 gain matrix. This state estimate combines an estimate from the internal model of the system dynamics together with a sensory correction modulated by the Kalman gain matrix $K(t)$. 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, $\beta = 3.9$ N·s/m, $m = 4$ kg and u were chosen based on the mass of the arm and the observed relationship between time and distance traveled. Specifically, the total force u was chosen to be linearly related to the average velocity under each of the three force conditions: 1.3, 1.5 and 1.9 N corresponding to the average movement velocities of 10.8, 12.8 and 16.6 cm s^{-1} for the resistive, null and assistive conditions respectively. To end the movement the sign of the motor command u_{int} was reversed until the arm was stationary. To simulate the overestimation of distance traveled $\hat{B}$ was set to $\begin{bmatrix} 0 \\ 1.4/m \end{bmatrix}$ while both $\hat{A}$ and $\hat{C}$ accurately reflected the true system. Noise covariance matrices of $Q = 9.5 \times 10^{-5} I$ and $R = 3.3 \times 10^{-4} I$ were used representing a standard deviation of 1.0 cm for the position output noise and 1.8 cm s^{-1} for the position component of the state noise.

Part II

Coordinate Transformations

Chapter 4

Representation of the Visuomotor Coordinate Transformation

4.1 Introduction

The human central nervous system (CNS) receives sensory inputs from a multitude of modalities, each tuned to extract different forms of information from the environment. These sensory signals are initially represented in disparate coordinate systems—for example visual information is represented retinotopically whereas auditory information is represented tonotopically. The ability to transform information between coordinate systems is necessary for both perception and action. In perception, coordinate transformations are required to convert sensory data into a common representational format so that they can be fused into a single percept. For example, visual and auditory stimuli arising from a common source can be combined into a single representation of the location of the source. In action, coordinate transformations are used to convert sensory information into coordinates appropriate for movement. For example, when we reach to a visually perceived object in space, the location of the object in visual coordinates must be converted into a representation appropriate for movement, such as the configuration of the arm required to reach the object. The coordinate transformation between the visual location, initially represented as a retinotopic pattern of neural activity, to the arm configuration required to place the hand at that location is known as the visuomotor map, and is the focus of the first part of this chapter. The coordinate transformation that maps the location of visual and auditory stimuli into a common reference frame is known as the visuo-auditory map, and is the focus of the second part of this chapter.

4.2 The Visuomotor Coordinate Transformation

Although the relationship between the visual and motor coordinate systems changes over time, due to factors such as growth, the visuomotor map can adapt to these changes. By examining the change in visuomotor coordination under prismatically induced displacement and rotation, Helmholtz (1867/1925) and Stratton (1897a, 1897b) pioneered the systematic study of this coordinate transformation. Their studies demonstrated both the fine balance between the visual and motor coordinate systems, which is disrupted by such perturbations, and the ability of subjects to adapt to the displacements induced by the prisms. Subsequently, many studies have further demonstrated the remarkable ability of subjects to adapt, at least partially, to a wide variety of alterations in the relationship between visual and motor system (for reviews see Welch, 1978 and Howard, 1982)—the single prerequisite for adaptation seems to be that the remapping be stable (Welch, 1986).

Two classes of hypotheses have been proposed to explain the mechanism for adaptation—sensory hypotheses (e.g. Harris, 1965) and sensorimotor hypotheses such as the reafference hypothesis of Held and colleagues (Held, 1962; Held and Hein, 1958). The sensory hypotheses propose that visuomotor adaptation is driven by changes in the normal relationship between vision and proprioception and that calibration is, therefore, between these two sensory information sources. At each point in time the subject both sees and feels the position of his or her hand and learns to correlate the two signals. The sensorimotor reafference hypothesis proposes that adaptation is driven by discrepancies between vision and the efference copy of the motor command. The visual location of the hand (reafference signal) is correlated with the motor command (efference copy) and a map is therefore formed between the sensory input and the motor output. In both sensory and sensorimotor theories adaptation takes place by simultaneously monitoring the visual position of the hand and a signal effectively coding—either through proprioception or motor efference—the arm configuration. Therefore, both these theories rely on a process of mapping or correlation between visual and arm configuration signals for calibration. While the conditions (e.g. Held and Hein, 1958), components (e.g. Harris, 1965; Redding and Wallace, 1988) and time course of adaptation (e.g. Dewar, 1970) have been extensively characterized, less is known about the topological properties of the visuomotor map. In this paper we examine the topological structure of the visuomotor map by introducing localized perturbations into the map and studying the resulting patterns of spatial and contextual generalization.

4.2.1 Spatial Generalization

One way in which the representation of the visuomotor map can be studied is through an examination of spatial generalization. That is, how does pointing change throughout the reaching workspace after exposure to a highly localized set of remapped

points? Consider a subject moving his arm while wearing prisms, but having the visual feedback of his arm limited in such a way that he receives concurrent visual and motor information only at a single point. Such a local remapping, which perturbs the visuomotor map at only a single location, is consistent with a wide variety of possible global remappings. The particular remapping that is chosen by the visuomotor control system, as demonstrated by the change in pointing behavior at different positions in the workspace, reflects intrinsic properties of the map (Bedford, 1989).

A coordinate transformation such as the visuomotor map can be regarded as a function relating one set of variables (inputs) to another (outputs). For the visuomotor map the inputs are the visual coordinates of a desired target and the outputs could be the corresponding motor coordinates representing the arm's configuration (e.g. joint angles). The problem of learning a sensorimotor remapping can then be regarded as a function approximation problem (Koh and Meyer, 1991). In function approximation, there is an explicit correspondence between the representation used and the patterns of generalization that will emerge. These function approximators can span patterns of generalization from local (look-up tables), through intermediate (CMACs, Albus, 1975; and radial basis functions, Broomhead and Lowe, 1988) to global (parametric models). We will return to a fuller discussion of these computational models in light of our data.

Several research groups have recently addressed the issue of generalization in visuomotor learning (Bedford, 1989; Imamizu et al., 1994). Bedford (1989) used a procedure in which subjects pointed to lit targets in the dark while looking through a prism. By controlling the illumination of a light emitting diode (LED), mounted on the subject's fingertip, Bedford was able to limit the subject's exposure to a single pairing of visually and proprioceptively felt finger position. Through the use of the prism, a discrepancy between vision and proprioception was induced such that the subject would feel the finger position at one location but see it at another. By comparing pointing behavior to a series of targets in an arc before and after training, Bedford assessed the change in the visuomotor map—that is, the extent to which learning a remapping at one point generalized to other points in space. Bedford found that training at one location generalized to the entire arc such that pointing shifted everywhere by the same amount. In further experiments subjects were trained on a remapping at two and three points and the change in pointing was again assessed throughout the arc. The results indicated that training at two points generalized linearly, that is, the change in pointing was a linear function of target position, and that training at three points, even when the remapping at the three points was not fully consistent with a linear remapping, again generalized linearly. Bedford concluded that learning between perceptual dimensions was constrained to generalize linearly (Bedford, 1989). In the current study, we report results for a two-dimensional pointing task. Moreover, we make use of an experimental apparatus in which targets and reaching movements are in the same physical locations in space (in Bedford's experiments, the targets were outside the reaching workspace and the subject was asked

to place their finger so that it lay in a plane containing the target and the subject's eye).

Bedford's study examined the visuomotor transformation along a single dimension. As the subjects were tested in one dimension, along an arc centered around the subjects' eyes, these results cannot provide a full picture of how the visuomotor transformation is represented. Thus, for example, Bedford's results do not distinguish between transformations of the kinematic map such as translation and rotation.

4.2.2 Contextual Generalization

Another way in which the representation of the visuomotor map can be investigated is by examining its behavior when confronted with multiple remappings of the same point in visual space. Several studies have shown that when different perturbations are separated spatially, subjects are eventually able to adapt to each perturbation in the appropriate part of space. For example, Kohler (1950) fitted a subject with half-prism spectacles, in which the upper half of the visual field was displaced by 10° and the lower half of the visual field was undisplaced. After a month's exposure the subject adapted to both upper and lower field displacements. Similarly, Shelhamer et al. (1991) have recently examined adaptation to multiple eye-position-dependent gains in the vestibulo-ocular reflex (VOR). Magnifying and minifying lenses were used to produce different amounts of retinal slip experienced per degree of head rotation depending on whether the eyes were looking up or down. The VOR is normally finely tuned to produce an eye movement opposite in direction and equal in magnitude to an experienced head rotation so that the visual image appears stable on the retina. Exposure to these lenses produces eye-position-dependent changes in the gain of the VOR, suggesting that eye position can modulate the VOR gain.

Other studies have examined how the visuomotor map responds to multiple mappings of the same visual location separated by time. For example, McGonigle and Flook (1978) studied prismatic adaptation over ten sessions with three day intervals between the sessions. Each session comprised of three sequential conditions—prismatically-induced leftward and rightward displacements and no prism deviation. An overall improvement to both prisms over the sessions, with greater improvement in earlier sessions, was found. Similarly, Welch et al. (1993) exposed subjects to alternating 15 diopter left and 15 diopter right prisms over 12 sessions, resulting both in an improved relearning of each displacement and an improved general ability to learn new displacements, such as one imposed by a 30 diopter prism. These studies demonstrate that repeated exposure to multiple remappings of a single point in space improve subjects' ability to readapt to each displacement.

Conditioned or *contextual* adaptation combines attributes of both of the above forms of adaptation. In this paradigm, distinct remappings of the same location in space can be elicited by experimentally manipulating a context variable. Previous contextual adaptation studies have shown that subjects elicit aftereffects de-

pendent on the feel of the prism goggles (Kravitz, 1972; Welch, 1971), an auditory tone (Kravitz and Yaffe, 1972), and the felt direction of gaze (Hay and Pick, 1966). An issue which has not been explored is how adaptation generalizes as the context is continuously varied. We explore this issue using a novel and natural context for the visuomotor map, the movement starting location and address whether multiple starting point dependent visuomotor maps can be concurrently represented and appropriately indexed. If such maps can be learned, then their representation can be probed by examining the modulating influence of the context on the visuomotor map. In other words, how does the visuomotor map generalize to novel contexts for the movement?

4.2.3 Experimental Aims and Overview

In the present study we have first sought to test how the visuomotor coordinate transformation changes in a two-dimensional workspace after remapping at only one and two input-output pairs (Experiments 1 and 2, respectively). In a second series of studies we examine generalization across context-dependent modules by remapping a single visual position to two different finger locations dependent on the location of the start of the movement (Experiments 3 and 4).

Many previous studies have investigated the effects of altered visual feedback using either optical devices such as prisms or visual feedback on a computer monitor separate from actual hand position (e.g. Cunningham, 1989). We could not easily use an optical system, such as a prism, to perturb the visual feedback of the actual arm as it was necessary for our experiments to have rapid position-dependent control of the nature and direction of the perturbation. On the other hand, the use of a computer monitor requires the subjects to make additional coordinate transformations to link their hand position to the cursor spot. We have, therefore, designed a two-dimensional virtual visual feedback apparatus in which the need for any coordinate transformation between cursor and hand position is obviated; the virtual image of the cursor is at the same position, in three-dimensional space, as the finger. Using this setup, described in more detail below, complex state-dependent perturbations can also be introduced. As Held et al. (1966) have shown, also using a virtual image setup, the use of a luminous spot to represent finger position is sufficient to elicit prismatic adaptation provided that the cursor spot and hand movements are tightly correlated.

4.3 Experiment 1: Visuomotor Generalization to a One-Point Displacement

In order to study the topology of the visuomotor map we have extended Bedford's method to a two-dimensional workspace. The aim of our first experiment was to assess the adaptation in the visuomotor map, as measured by the change in pointing behavior, after a period of exposure to a single remapped input-output pair. This was achieved by restricting the visual feedback of the subject's finger, as represented by a cursor spot, to within a few millimeters of the remapped point. When the subject was outside this area the cursor spot was extinguished. Before and after this exposure phase the subject's pattern of pointing was assessed to locations on a grid of nine targets. These movements were performed in the dark. In distinction to Bedford's study, where subjects were asked to "point so that it feels like your right eye, the tip of your finger, and the light in space are lined up" (Bedford, 1993), we instructed subjects to place their finger as accurately as possible at the exact location where they saw each target, thus matching both direction and distance.

The change in pointing behavior in the absence of visual feedback is a measure of the prismatic aftereffect. Thus, one might expect to see some adaptation at the training point. However, as the subject was given no information about the mapping at any location other than the training point, any change in pointing behavior at the other targets is an indirect effect of the training. As the task places no constraints at the non-training targets, the pattern of generalization obtained is a result of intrinsic constraints on the representation of the visuomotor mapping (Bedford, 1989, 1993).

Several possible patterns of generalization arise from different hypotheses (Figure 4-1). If the limited exposure at the central training point is insufficient to produce any adaptive effect then no change in pointing behavior would be expected. However, if the mapping is represented locally, training at one point might result in aftereffects at the training point alone. Alternatively, if the mapping is represented globally then some form of adaptation would be expected at the outer eight targets. For example, if the central remapping was interpreted as a change in felt direction of gaze, a common finding in the prism adaptation literature (e.g. Welch, 1986 for a review), the pattern of generalization might be a rotational change. Alternatively, as suggested by Bedford's studies, the visuomotor map may be constrained to generalize linearly. For example, a linear translation in Cartesian coordinates might be seen, which would result in a pattern of aftereffects very different from that arising from laterally displacing prisms, whose primary effect is to induce a fixed angular rotation of the visual field. Another possibility is that generalization may be Cartesian but the effect may decay with distance from the training point. Lastly, some other non-linear pattern of generalization may result consistent with alternate possibilities such as joint- or muscle-based adaptation.

4.3.1 Method

Subjects

24 right-handed undergraduate students participated as subjects. Subjects were naive to the purpose of the experiment and were paid $7.00 for participation. All subjects had self-reported normal or corrected-to-normal vision.

Apparatus

In order to measure pointing behavior and to constrain subjects to experience limited input-output remappings we designed a two-dimensional virtual visual feedback setup. This consisted of a digitizing tablet to record the finger position on-line and a projection/mirror system to generate a cursor spot image representing the finger position. This setup allowed us to project the virtual image of the finger as well as targets in the plane of the table. The exact relation between the cursor spot and finger position could be controlled on-line so as to generate alterations in the visuomotor map. Furthermore, the cursor spot could be illuminated and extinguished so as to allow concurrent visual-proprioceptive feedback in restricted areas of the workspace. This setup is described in more detail below.

Subjects sat at a large horizontal digitizing tablet (Super L II series, GTCO, MD) with their head supported by a chin and forehead rest (Figure 4-2). This placed the subjects' eyes in a plane approximately 25 cm above the digitizing tablet. The subject's right index finger was mounted on the cross hairs of a digitizing mouse which could be moved along the surface of the digitizing tablet; the subject's arm was hidden from direct view by a screen. The digitizing tablet's coordinates were sampled as (x, y) coordinate pairs at 185 Hz by a PC; the accuracy of the board was 0.25 mm.

The targets and the feedback of finger position were presented as virtual images in the plane of the digitizing tablet (and therefore in the plane of the finger tip). This was achieved by projecting a Video Graphics Array (VGA) screen (640 x 480 pixels) with an LCD projector (Sayett Media Show) onto a horizontal rear projection screen suspended 26 cm above the tablet (Figure 4-2). One pixel measured 1.2 x 1.2 mm on the screen. A horizontal front-reflecting semi-silvered mirror was placed face up 13 cm above the tablet. The subjects viewed the reflected image of the rear projection screen by looking down at the mirror. By matching the screen-mirror distance to the mirror-tablet distance all projected images appeared to be in the plane of the finger when viewed in the mirror. Targets were presented as 9×9 pixel (10.8 mm) hollow squares and the finger position was indicated by a 5×5 pixel (6 mm) filled white square (cursor spot). The position of the finger was used on-line to update the position of this cursor spot at 50 Hz.

Prior to each experiment the position of the digitizing mouse cross-hairs relative to projected pixel position was calibrated over a grid of 16 points on the tablet. By illuminating the semi-silvered mirror from below, the virtual image and the cross-

hairs of the digitizing mouse could be lined up by eye. A quadratic regression of x and y pixel position on x and y hand position was performed and this was used on-line to position the targets and cursor spot. The correlation of the fit was always greater than 0.99. Cross-validation sets gave a average calibration error of 1.5 mm.

Procedure

Subjects were randomly assigned to one of three groups: control, x-shift and y-shift. Each experimental session consisted of four parts.

In the first part (familiarization phase) the subject was familiarized with the setup by pointing eight times to each of nine randomly presented targets on a 3×3 grid. Pointing movements were made under full visual feedback of finger position, as represented by the cursor spot. The target appeared and remained illuminated until the subject moved the cursor to the target position. The target then disappeared and the next target appeared when the subject had moved at least 15 cm away from the previous target.

In the second part (pre-exposure phase), the subject's pointing accuracy was assessed in the absence of visual feedback of finger position. The subject was instructed to point as accurately as possible to visually presented targets. The subjects indicated when they thought their finger was on target by pressing a mouse key with their left hand. Subjects were encouraged to be as accurate as possible and to press the mouse key only when they thought their finger position matched the target exactly. The target then disappeared and the next target appeared when the subject had moved 15 cm away from the previous target. This ensured that relative direction of the targets could not directly cue the subject's pointing movement. Targets were presented eight times each in a pseudorandom order on the same 3×3 grid. The subjects received no information as to their pointing performance. During this phase the target and finger positions were recorded for each trial.

The third part (exposure phase) of the experiment was designed to provide extensive exposure to an altered mapping between the visual and proprioceptive systems at a single location at the center of the workspace. Subjects were instructed to point to a central visually presented target—the training point. The cursor spot representing their finger position was only illuminated when it was within 0.5 cm of the target box. This allowed only very limited concurrent visual-motor feedback.

The relationship between the cursor spot and actual finger position was altered for the different groups. For the control group the finger cursor accurately represented the finger position. Therefore in order to see the cursor on target their finger had also to be on target. For the other two groups a discrepancy was introduced between the actual and perceived finger position (Figure 4-3a). For the x-shift group the subject had to point 10 cm to the right of the central target in order to see the cursor spot on target (Figure 4-3b). For the y-shift group the subjects had to point 10 cm towards their body in order to see the cursor spot on target (Figure 4-3c). In these two groups the subjects were, therefore, exposed to a single remapping of finger position to visual

position. Once the central target was reached the subject had to maintain the finger cursor there for 2 seconds, until the target turned from white to blue and one of the 8 peripheral targets became illuminated in a pseudorandom order. The subject then had to move towards that target; after having moved 15 cm the central target would turn white and the cycle would repeat. The subject pointed a total of 40 times to the central target.[1]

Limiting the area of the cursor feedback to within 0.5 cm of the target ensured that during the exposure phase the visual and proprioceptive information on finger position co-occurred only within a small region of the workspace. However, this limited feedback also made the task of pointing to the central target difficult. Subjects were warned that this phase of the the experiment would be difficult and that they would have to try moving around to find the target. To aid the subject in finding the target, after 10 seconds, one of the following messages would be displayed at the bottom of the screen—"try left", "try up", "try right" or "try down". A random search strategy such as Bedford's, where subjects were told "try moving your hand back and forth slowly" (Bedford, 1989) could not be employed since in a two-dimensional workspace it is not guaranteed to locate the target. During this exposure phase the time to place the finger on target was recorded as a measure of visuomotor learning of the training target.

The final phase (post-exposure phase) was identical in form to the second (pre-exposure) phase; subjects' pointing was again measured, in the absence of cursor feedback, on the 3×3 grid with eight repetitions at each point. The pseudorandom order of the targets was changed from the second phase.

For the control and x-shift groups the grid points were evenly spaced on a square from (-10,20) to (20,50) cm relative to the midpoint between the eyes (Figure 4-3a). For the y-shift group the grid was reduced evenly in the y-direction by 10 cm to (-10,25) to (20,45) cm. This was necessary because if the subject adapted fully to the 10 cm perturbation, the closer target points would be reached with movements outside the recording area of the tablet. In all cases the position of the central target was maintained at (5, 35) cm.

Analysis

To study the effect of initial pointing inaccuracies the pre-exposure pointing errors were analyzed in each group separately. The average finger position for each target was calculated together with its covariance matrix. The average pointing locations

[1]Due to an error in experimental coding, the control sessions were slightly shorter, requiring only 30 repetitions of the central target pointing cycle in the exposure phase. However, in both this experiment and in Experiment 2, where the control consisted to 60 repetitions of a similar cycle with no perturbation, the control groups exhibited almost no change in pointing between pre- and post-exposure phases. This indicates that the basic premise of the experimental control, that subjects' pointing does not change systematically during an unperturbed exposure phase, is valid.

were plotted, together with their corresponding targets, as 95% confidence ellipses centered around the sample mean.

To assess the improvement in target acquisition during the exposure phase the mean time to reach the target over batches of five trials was plotted. To assess generalization of the visuomotor map, the subjects' change in pointing behavior between the pre-exposure and post-exposure phases was analyzed. For each subject and target the average change in pointing position between the pre-exposure and post-exposure phases was calculated, along with the corresponding covariance matrices. The subjects' data was combined within each group and target, yielding the average change for the group and the covariance matrix for each target. Each vector change and covariance matrix is based on 128 data points (8 subjects $\times$ 8 repetitions $\times$ pre- and post-exposure conditions). The mean change in pointing position for each target was plotted at that target as an arrow along with the 95% confidence ellipse. These plots, therefore, show the change in the pointing behavior subsequent to the exposure phase while factoring out any consistent inaccuracies in pointing.

Per target analyses (ANOVAs) of x and y pointing errors were performed to assess the significance of the change in pointing at each target, with phase as the within-subject factor. The significance of the overall changes in pointing errors was assessed through separate ANOVAs for each group, with phase (pre- and post-exposure) and target (9 locations) as within-subject factors.

Two alternative representations were also used to display the data. First, an interpolated vector field from the mean change vectors was obtained by Gaussian kernel smoothing (kernel width s.d. 7.0 cm). The Gaussian kernel smoothed fields were also used to estimate the proportion adaptation in the direction of the perturbation, which were plotted as greyscale contour plots. These contour plots, therefore, display an estimate of the proportion adaptation over the workspace.

4.3.2 Results

Pre-exposure errors

Subjects showed a consistent pattern of pointing errors in the pre-exposure phase. The pattern of inaccuracies in initial pointing was similar between groups and generally showed a bias away and towards the left side of the subject (Figure 4-4). In particular, pointing at the central training point was biased away and to the left for all three groups.

Learning during the exposure phase

During the exposure phase, due to the limited feedback, the target was difficult to find. Figure 4-5 shows how the time to acquire the target changed as a function of practice. For both the x and y shift groups the target took initially longer to acquire than in the control. Over the course of the exposure phase, the time to acquire the

Experiment 1: Analysis of Variance							
Group	Dir	Phase		Target		Phase × Target	
		$F_{1,7}$	p	$F_{8,56}$	p	$F_{8,56}$	p
Control							
	x	2.16	ns	3.75	< 0.01	< 1	ns
	y	< 1	ns	2.57	< 0.05	1.61	ns
X-Shift							
	x	15.83	< 0.01	5.96	< 0.001	2.06	ns
	y	1.96	ns	9.51	< 0.001	1.36	ns
Y-Shift							
	x	< 1	ns	1.35	ns	2.11	< 0.05
	y	3.75	ns	< 1	ns	1.87	ns

Table 4.1: Summary of the two-factor within-subject ANOVAs for the three experimental groups and two directions (Dir) in Experiment 1. Non-significant effects at the $\alpha = 0.05$ level are denoted by ns.

targets dropped to levels not significantly different from the controls.

Generalization

The pattern of generalization for the controls is shown in Figure 4-6a. The figure represents the change in pointing between pre- and post-exposure phases plotted as vectors centered at the 9 targets. For example, a 1 cm leftward-pointing arrow would signify that subjects' pointing to that target changed by 1 cm to the left between the pre- and post-exposure sessions. The ellipses centered at the arrow tip are 95% confidence ellipses for the change in the sample mean. The per target ANOVAs reveal that none of these changes are significant at the $\alpha = 0.05$ level. The interpolated vector field of changes for the control (Figure 4-6b) highlights the fact that, although there were no significant changes, there was a small trend towards the left for all 9 points.

The ANOVA (summarized in Table 4.1) shows no significant main affect of phase for the x or y directions. The main effect of phase indicates the global component of change between the pre- and post-exposure phases. Therefore, the control subjects, as expected, did not change their pointing behavior in either the x or the y direction.

We now consider the effect of introducing a remapping at one input-output pair. The general effect of introducing such a perturbation was to induce significant changes in the pointing behavior not only at the remapped point but at neighboring points as well. The pattern of generalization for the x-shift group is shown in Figure 4-7a; the change in pointing between the pre- and post-exposure phases was significant at 6 out of the 9 targets (targets 1–6, numbered according to Figure 4-3b) in the x direction

and at 1 out of 9 targets (target 9) in the y direction. The shift was greatest at the training point (4.9 cm) and decreased in magnitude away from this point. The overall ANOVA (summarized in Table 4.1) shows a significant main effect of phase for the x direction indicating a global change between the pre- and post-exposure phases.

The interpolated vector field of changes for the x-shift group (Figure 4-7b) shows a pattern of decaying rightward changes with a downward y trend further from the subject. The proportion adaptation in the direction of the perturbation computed from the vector fields is depicted in Figure 4-7c as a greyscale contour plot. This shows the pattern of greatest change occurs at the training point and decays with distance away from it.

The pattern of generalization for the y-shift group is shown in Figure 4-8a. The change in pointing between the pre- and post-exposure phases was significant at 1 out of 9 targets (target 8) in the x direction and at 3 out of 9 targets (targets 1, 2 and 5) in the y direction. The change in the y direction at target 8 was marginally significant ($p = 0.06$). As in the x-shift group, the shift was again greatest at the training point (2.2 cm). Changes were most pronounced at the two rows closest to the subject; there were no significant changes in the row of targets furthest from the subject. The overall ANOVA indicates that the y direction of change in the y shift group was marginally significant ($F_{1,7} = 3.75, p = 0.09$).

The interpolated vector field of changes for the y-shift group is shown in Figure 4-8b. This highlights the pattern of downward (i.e. towards the body) changes decaying away from the training point. The proportion adaptation contour plot (Figure 4-8c) again highlights a pattern of adaptation that is greatest near the training point and decays away from it.

4.3.3 Discussion

The learning curves indicate that control subjects were initially better than either perturbation group at locating the training point during the exposure phase. This is to be expected, since their visuomotor map was unperturbed. The x-shift group took slightly longer to locate the target and the y-shift group took considerably longer. Taking time to locate the target as a measure of task difficulty this suggests that the y perturbation, through identical in magnitude, was more difficult to adapt to than the x perturbation. This is also supported by the observation that the proportion adaptation for the y-shift group was smaller than for the x-shift group. After about 30 pointing trials for both the x-shift and y-shift groups the time to attain the target decreased to control levels.

Although there was a large consistent pattern of pre-exposure errors, repeated unperturbed training at the central target in the control did not decrease these errors. This indicates that the pattern of pre-exposure pointing errors does not seem to be corrected for with training at the central target.

The effect of remapping a single point in the visuomotor map was to induce

significant global changes in the pointing behavior. It is important to note that as subjects were both uninformed and unaware of the perturbation the adaptation can be regarded as perceptual rather than cognitive (Bedford, 1993). For the x-shift group the large shift in the compensatory direction at the training point indicates a substantial local aftereffect (47%) due to the exposure. Five out of eight peripheral locations also showed a significant shift in the compensatory direction—a finding that is is inconsistent with a purely local model of the adaptation process. When viewed in the contour plot in Figure 4-7c the global effect can be seen as a decaying surface with its peak centered at the training point.

Similarly, a remapping in the y-shift direction appears to induce the largest shift in the compensatory direction at the training point, showing a local aftereffect of 21%. The pattern of peripheral shifts is somewhat different than in the x-shift group: three of the five peripheral targets nearest to the body showed a significant or near-significant y shift, while none of the three furthest targets showed a shift approaching significance. This suggests that the effect is not local but that the global effect decreases further away from the body (Figure 4-8c).

Taken together, the y-shift and x-shift data are not consistent with a model that represents the learning as a change in felt direction of gaze (Harris, 1965). Due to the arrangement of the chinrest and table, the subjects' eyes are sagittally away from (35 cm) and above (25 cm) the position of the training point. If the adaptation were represented as a constant angular offset in the felt direction of gaze one would have expected larger shifts in pointing at the more distant targets for both the y-shift and x-shift groups—in fact these shifts were generally smaller.

In summary, the one-point generalization study shows that the generalization appears to be global but that the effect falls off with distance from the exposure site. The results suggest that training at one point has a non-linear generalization effect; that is, the changes are not uniform across the workspace. Based on this finding one would expect that training at two points might result in a pattern of generalization not consistent with Bedford's linear constraint hypothesis (Bedford, 1989, 1993). To explore this question further and to further elucidate the constraints on the map we conducted a two point generalization study.

4.4 Experiment 2: Visuomotor Generalization to a Two-Point Displacement

In Experiment 2 subjects were exposed to perturbations of the normal relation between vision and proprioception at two points. The experimental question was again, how does pointing behavior change after exposure to local perturbations in the visuomotor map? We chose the perturbations at the two points to be of opposite sign in the y (sagittal) direction to test the hypothesis that the map was constrained to gen-

eralize linearly. Such a perturbation, displayed in Figure 4-9, introduces a conflict if the map were to be interpreted in a globally linear way. That is, the Cartesian linear hypothesis (Figure 4-1) would predict for each perturbation a globally linear generalization of opposite sign, thereby cancelling to produce no generalization. On the other hand, the Cartesian decaying hypothesis suggested by Experiment 1 predicts that the two perturbations will each generalize to the region of space around them. However, there are many other possible patterns of generalization consistent with the perturbation; for example, a counterclockwise rotation about the central target or a skew transform. Both of these patterns of generalization are linear transformations of Cartesian space (see discussion of possible transformations in Bedford, 1993).

4.4.1 Method

Subjects

16 naive right-handed undergraduate students participated in this study. Subjects were paid $7.00 for participation. All subjects had self-reported normal or corrected-to-normal vision.

Procedure

Subjects were randomly assigned to one of two groups: control and y-shift. The paradigm was identical to the one-point generalization experiment except that in the pre- and post-exposure phases 11 points were tested and in the exposure phase training alternated between two targets. These differences are detailed below.

In the pre- and post-exposure phases, subject's pointing accuracy was assessed in the absence of visual feedback of finger position at 11 targets (Figure 4-9). As in Experiment 1 pointing consisted of 8 pseudorandom repetitions at each target. Nine of the targets were identical in location to those used in Experiment 1. The other two targets were located to the left and the right of the central target and were used as training points during the exposure phase.

The workspace used in Experiment 2 was identical to that used for the control and x-shift groups in Experiment 1. Based on Experiment 1, we realized that subjects did not generally adapt fully to the 10 cm perturbation, and therefore it was unnecessary to reduce the workspace as was done for the y-shift group in Experiment 1.

During the exposure phase of this experiment, two training locations were used: one on the left (-2.5, 35.0) and one on the right (12.5, 35.0) of the grid center (targets 10 and 11 in Figure 4-9). The paradigm was similar to the one point study except that subjects alternated between pointing to the left and right target for a total of 60 repetitions, 30 repetitions at each target. For the control group the cursor accurately represented finger position. For the y-shift group the subject had to point 10 cm towards the body at the left target and 10 cm away form the body at the right target so as to appear on target (arrows in Figure 4-9).

Analysis

The analysis was identical to that performed in Experiment 1, except for the increased number of targets. To obtain the interpolated vector fields and contour plots the Gaussian kernel width of the smoothing algorithm was reduced to 3.5 cm, since there was, in this experiment, a higher density of data points collected over the same workspace. The time to reach the target was batched over 10 trials.

4.4.2 Results

Pre-exposure errors

Subjects showed a consistent pattern of pointing errors in pre-exposure phase (Figure 4-10). The pattern of inaccuracies in initial pointing were similar to those found in Experiment 1 (Figure 4-4). In particular, the pattern of pre-exposure pointing errors for the y-shift group displayed the same tendencies of overall overshoot, larger for the three targets on the right, with a leftward bias for the targets on the left. For the control group, overshoot was only present in the distant targets, with significant undershoot in the three near targets.

Learning during exposure phase

Figure 4-11 shows how the time to acquire the target changed as a function of practice. Initially, the target took significantly more time to acquire for the y-shift group than for the control. Over the course of the exposure phase the time to acquire the targets dropped to levels not significantly different from the controls.

Generalization

The pattern of generalization and interpolated vector field for the control group are shown in Figure 4-12; the per target ANOVAs indicated that none of the changes in pointing were significantly different from zero. The ANOVA, summarized in Table 4.2, shows no significant main effect of phase for the x or y directions, indicating that the control subjects did not change their global pointing behavior in either the x or the y direction.

Figure 4-13a shows the pattern of generalization for the y-shift group. The change in pointing between the pre- and post-exposure phases was significant at 2 out of 11 targets (targets 3 and 6) in the x direction and at 4 out of 11 targets (targets 8–11) in the y direction. Additional marginally significant ($p < 0.10$) changes occurred at 1 target (targets 4) in the x direction and at 4 out of 11 targets (targets 1, 2, 4 and 6) in the y direction. The change was greatest at the right training point (6.2 cm), followed by the target immediately to its right (4.9 cm), and then at the left training point (4.7 cm).

Experiment 2: Analysis of Variance							
Group	Dir	Phase		Target		Phase × Target	
		$F_{1,7}$	p	$F_{10,70}$	p	$F_{10,70}$	p
Control							
	x	< 1	ns	6.73	< 0.0001	2.22	< 0.05
	y	< 1	ns	9.73	< 0.0001	1.54	ns
Y Shift							
	x	< 1	ns	12.66	< 0.0001	1.82	ns
	y	< 1	ns	16.54	< 0.0001	12.73	< 0.0001

Table 4.2: Summary of the two-factor within-subject ANOVAs for the three experimental groups and two directions (Dir) in Experiment 2. Non-significant effects at the $\alpha = 0.05$ level are denoted by ns.

The pattern of generalization in the interpolated vector field of changes for the y-shift group (Figure 4-13b) shows a change in pointing away from the body in the upper right half of the workspace and towards the body in the lower left half. The ANOVA (Table 4.2) showed no significant main effects of phase but a highly significant interaction of phase and target in the y direction, reflecting the non-linear effect.

The proportion adaptation in the direction of the perturbation (y direction) computed from the vector field is depicted in Figure 4-13c as a greyscale contour plot. The lighter areas represent change away from the body and the darker areas change towards the body.

4.4.3 Discussion

Again, although there were large consistent pre-exposure pointing errors the control only displayed small changes between the pre- and post-exposure phases. This reinforces the notion that the visuomotor map is relatively stable throughout the experiment when not exposed to a perturbation.

The learning curves for the exposure phase suggest that the two-point perturbation was initially more difficult to compensate for than either one-point perturbation from Experiment 1. However, as the session proceeded the time to locate the two targets decreased to control levels.

The effect of perturbing the visuomotor mapping at two points in the workspace was to induce a pattern of generalization around each of the points which, as in the first experiment, decayed away from the training points. As the perturbations were opposite in direction, there was a region between the two training points where the visuomotor map did not change. The results of this experiment suggest that the effect of simultaneously remapping several points can be explained qualitatively as

an effect of superimposing the contributions of several single point remappings. The perturbation at two points is not interpreted by the visuomotor map as a single global remapping. In particular, the two-point perturbation could have been interpreted by the visuomotor system as a single global remapping consisting of a counterclockwise rotation about the central target. This would have produced a pattern of generalization with large opposite-sign x-shifts at the middle-top and middle-bottom targets. However, neither these nor the other peripheral targets demonstrate the pattern of changes predicted by a rotatory shift.

The results of the two-point perturbation study are at odds with the conclusions that Bedford draws from her two-point perturbation results. Experiments 2 and 3 of Bedford (1989) examined generalization to two-point perturbations using lateral displacements at two training targets along an arc centered about the subjects' eyes. Bedford examined the change in pointing behavior before and after this perturbation at three positions between the two training points and at six positions outside the range of the training points. Subjects displayed neither significantly larger nor significantly smaller shifts in the extrapolative region than at the training points. Our results from the two-point perturbation experiment also do not display larger shifts at peripheral targets than at the training points, at odds with a strict interpretation of Bedford's linear constraint hypothesis, although consistent with her data. Moreover, the results from Experiment 1 indicate that adaptation was consistently smaller away from the training point, rather than staying constant or growing linearly. Therefore, neither the results from the one-point nor the two-point perturbation studies are consistent with the linear constraint hypothesis.

Differences in methodology may explain the difference between Bedford's results and our results. In particular, our apparatus allowed a natural mapping between target locations and the subjects' movements: subjects were asked to position their finger at the same perceived location in space as the target. In Bedford's experiment, subjects were asked to place their finger in a vertical plane that also contained the target and their right eye. This is a more difficult task than ours, and the differences in difficulty may have influenced the patterns of generalization. Another possibility is that the differences are due to the two-dimensional nature of our task; in particular, our two-point perturbation involved a perturbation in depth (i.e., in radial distance from the subject), whereas Bedford's perturbations were all at constant depth.

In Experiment 1 and 2 we have examined patterns of spatial generalization in an attempt to elucidate the constraints underlying the representation of the visuomotor map. In the next series of experiments, we extend this paradigm to the study of contextual generalization. Two questions will be addressed: Can the same point in visual space be mapped onto two different hand locations depending on a context? If two separate context-dependent mappings can be induced, how does the visuomotor map generalize to other contexts?

4.5 Experiment 3: Contextual Generalization of the Visuomotor Map A

In this experiment we investigated the changes induced in the visuomotor map by a context-dependent remapping of a single visual location to two different finger positions. Subjects were exposed to different visuomotor rearrangements at a single visual target location during movements made from two possible starting locations. Two perturbations, equal in magnitude but opposite in sign, were used—where the sign of the perturbation was determined by this starting point of the movement. The start point of the movement, therefore, represented the context of the remapping. Such a remapping sets up both ambiguous and conflicting visuomotor pairs at this single visual location which can be resolved only through the use of the context.

The first goal of this study was to assess whether context-dependent maps can in fact be learned. Given that equal and opposite remappings of the same point are presented, one possible prediction is that there would be no change from either starting location; any potential adaptation due to one remapping would be counteracted by the other. Alternatively, if adaptation takes place it may be either context-independent or context-dependent. Context-independent adaptation would be found if the changes induced in the visuomotor mapping are the same from both starting locations, and conversely, context-dependent adaptation would be found if the changes are different from the two starting locations.

The second aim of this study was to explore how any changes induced in the visuomotor map generalize as the context is varied between the two starting locations. The use of starting position as the context for the movement allowed both a natural context for pointing movements and one in which the context could be continuously varied by selecting new start locations at points along the line joining the two initial starting points. If context-dependent adaptation is seen for the two start points used in training, then the form of generalization of this adaptation to new contexts can be assessed. As in the spatial generalization experiments, the extent of contextual generalization could range from none, through linear, to nonlinear. In particular, we were interested in assessing whether there would be an abrupt or smooth transition from one map to the other as the context is varied between the two learned contexts.

The paradigm in this experiment was similar in design to Experiments 1 and 2. Subjects now repeatedly pointed to a single visual target from several different starting points. This pointing behavior was assessed both before and after exposure to the remapping. During the exposure phase a perturbation was chosen so that a single visual location was remapped in one direction when approached from one starting point and another direction when approached from the other starting point. The perturbation was chosen to be equal and opposite in sign from the two starting points. The change in pointing behavior from different starting points was used to assess the context-dependent changes in the visuomotor map.

4.5.1 Method

Subjects

32 right-handed undergraduate students participated in this study. Subjects were naive to the purpose of the experiment and were paid $7.00 for participation. All subjects had self-reported normal or corrected-to-normal vision.

Apparatus

The apparatus was identical to that used in Experiments 1 and 2.

Procedure

Subjects were randomly assigned to one of four groups: control, open x-shift, crossed x-shift, and y-shift. Each experimental session consisted of four parts.

In the first part (familiarization phase) the subject was familiarized with the setup by pointing 3 times to each of 9 randomly presented targets on a 3×3 grid. The grid of targets was identical to that used in Experiment 1. Pointing movements were made under full visual feedback of finger position, as represented by the cursor spot. The target appeared and remained illuminated until the subject moved the cursor to the target position. The target then disappeared and the next target appeared when the subject had moved at least 15 cm away from the previous target.

In the second part (pre-exposure phase), the subject's pointing was assessed in the absence of visual feedback of finger position. The subject was instructed to point as accurately as possible to a single visual target located at (5, 40) cm relative to the midpoint of the subjects' eyes. The subject started each pointing movement from one of 7 possible starting locations arranged in a line at 7.5 cm intervals from (-17.5, 20) to (27.5, 20) cm (see Figure 4-14 for a layout of the starting points and the target). Starting points were selected in a pseudorandom order, 10 times each. At the beginning of each pointing trial, one of the seven starting points would be displayed as a hollow white square. The subject moved to this starting point, receiving finger feedback in the form of a cursor spot only when within 5 cm of this point. Having reached the starting point, its color turned from white to blue and the color of the target also changed from blue to white. The cursor spot was extinguished and subjects had to point to the target, indicating when they thought their finger was on target by pressing a mouse key with their left hand. Subjects were encouraged to be as accurate as possible and to press the mouse key only when they thought their finger position matched the target exactly. The subject received no information as to pointing performance. During this phase, the target and finger positions were recorded for each trial.

The third part (exposure phase) of the experiment was designed to provide extensive exposure to an altered mapping between visual locations and corresponding motor coordinates. Throughout this phase three hollow white 0.5 cm squares were

continuously displayed, corresponding to the left starting point, right starting point, and the target (Figure 4-14a). Note that the left starting point, right starting point, and target in the exposure phase correspond to starting points 2, 6, and the target, respectively, in the pre-exposure phase. Also displayed continuously was the cursor feedback of the subjects' finger position. One of the three boxes was always highlighted by changing its color from white to blue. The subject was instructed to always move to the box that had been highlighted. When the finger cursor reached the highlighted box the box would turn white and another box would be highlighted. The boxes were highlighted in a repeating sequence left-right-target-right-left-target—tracing out a triangle in which the subject would alternately point to the target from the left and right starting points. This sequence was repeated 40 times, such that the target was approached 40 times from each starting point.

The relationship between the cursor spot and actual finger position was altered for the different groups. For the control group, the finger cursor accurately represented the finger position. The subjects in the control group both traced out a triangle with their finger and also saw, veridically, a triangle being traced out visually (Figure 4-14a). For the three other groups a different displacement was introduced depending on the starting point and increasing linearly from starting point to target. Thus, for the crossed x-shift group the cursor spot was displaced by 5 cm to the left for movements from the left starting point and 5 cm to the right for movements from the right starting point. That is, while the subject visually perceived the finger tracing out a closed triangle the finger was actually tracing out a path that crossed at one point ending up on opposite sides of the target from the starting point (Figure 4-14b). For the open x-shift group the cursor spot was displaced at the target by 5 cm to the right for movements from the left starting point and 5 cm to the left for movements from the right starting point. These subjects again visually perceived their finger cursor moving along a closed triangle although, in this case, their actual finger moved along a opened triangle (Figure 4-14c). For the y-shift group the cursor spot was displaced by 5 cm away for movements from the left starting point and 5 cm towards the subject for movements from the right starting point (Figure 4-14d).

For all conditions there was no perturbation in movements made between the start points. For the two x-shift groups the displacement was a linear function of the distance traveled by the finger from the start point to the target along the y direction. Conversely, for the y-shift group the displacement was a linear function of the distance traveled along the x direction. This ensured that the displacement always started at zero and increased linearly to 5 cm at the target. The movements from the target back to the starting point had exactly the same displacement function ensuring a consistent perturbation in both directions of travel. The displacement varied smoothly from one extreme to the other while continuously displaying the finger cursor.

The final phase (post-exposure phase) was identical in form to the second phase (pre-exposure phase); subjects' pointing was again measured, in the absence of cursor feedback, 10 times from each of the 7 starting locations. The pseudorandom order of

the targets was changed from the second phase.

Analysis

To assess generalization to the context of the movement, the subjects' change in pointing behavior between the pre-exposure and post-exposure phases was analyzed for each starting location. For each subject and start location the average change in pointing position between the pre-exposure and post-exposure phases was calculated, along with the corresponding covariance matrices. The subjects' data were combined within each group for each starting target obtaining the group sample average change along with the covariance matrix for each start point. Each vector change and covariance matrix is based on 160 data points (8 subjects $\times$ 10 repetitions $\times$ pre- and post-exposure conditions). The mean change in pointing position from each starting point was plotted at that start location as an arrow along with the 95% confidence ellipse centered on the sample mean change. For each group the change in pointing in the direction of the perturbation was plotted as a function of the starting position.

The change in pointing as a function of starting location was analyzed through separate ANOVAs for each group, with phase (categorical pre- and post-exposure) and starting point (continuous x location) as within-subject factors.

4.5.2 Results

Although the perturbations were quite strange in nature—a single visual target being mapped onto two different finger positions and two different visual positions being mapped into the same finger position in the crossed x-shift group—subjects found the task simple and natural. Subjects were unaware that their finger feedback had been perturbed as revealed by informal questioning.

Figure 4-15 shows the pattern of changes in pointing to the target as a function of the starting position for the four groups. The change in pointing for the control group is shown in Figure 4-15a. The ANOVA shows no significant changes between the pre- and post- exposure phases (Table 4.3). Specifically, the lack of interaction of phase and target in both the x and y direction indicates that there was no significant linear trend in the change in pointing as a function of starting position. This is also confirmed in the control portions of the adaptation plots (Figures 4-16a & b, hollow circles).

Figure 4-15b shows the change in pointing for the crossed x-shift group. There is a small amount of overshoot in the y direction (mean 0.9 cm), a change to the left from the starting points on the right and a small change to the right for the leftmost target. The ANOVA shows a significant main effect of phase on x pointing position and a highly significant interaction of phase and target. The form of this interaction is revealed by the plot of x adaptation as a function of starting point (Figure 4-16a, solid circles). The perturbation is in the positive x direction from the starting point

Experiment 3: Analysis of Variance							
Group	Dir	Phase		SP		Phase x SP	
		$F_{1,7}$	p	$F_{1,7}$	p	$F_{1,7}$	p
Control							
	x	< 1	ns	1.41	ns	< 1	ns
	y	< 1	ns	8.49	ns	2.42	ns
Crossed x-shift							
	x	9.23	< 0.05	3.01	ns	33.54	< 0.001
	y	1.62	ns	8.26	< 0.05	< 1	ns
Open x-shift							
	x	2.17	ns	1.45	ns	2.45	ns
	y	9.00	< 0.05	3.11	ns	< 1	ns
y-shift							
	x	< 1	ns	< 1	ns	6.66	< 0.05
	y	< 1	ns	3.39	ns	23.3	< 0.01

Table 4.3: Summary of the two-factor within-subject ANOVAs for the four experimental groups and two directions (Dir) in Experiment 3. SP denotes starting point. Non-significant effects at the $\alpha = 0.05$ level are denoted by ns.

on the left and in the negative x direction from the starting point on the right. As the starting point is varied from left to right the adaptation displays a monotonic change in the x direction consistent with the direction of this perturbation.

For the open x-shift group there was a large change in pointing in the y direction (Figure 4-15c). After exposure to the perturbation subjects pointed further away from the target by 2.8 cm on average. This pattern of overshoot was seen in 7 out of 8 of the subjects. The ANOVA shows this significant effect of phase on y pointing position but no significant effect on x. In particular, there was no interaction of phase and starting point on x pointing position as would be predicted by adaptation to the perturbation. Comparison of the pattern of changes in the x direction between this group and the control (as plotted in Figure 4-16a) with a between-groups ANOVA revealed no significant differences between the two groups.

The pattern of changes for the y-shift group is shown in Figure 4-15d. While there are no significant main effects of phase there is a interaction of phase and starting point on both x and y pointing position (Table 4.3). In particular, Figure 4-16b shows the y adaptation as a function of starting point as compared to the control. For this group, the perturbation is in the negative y direction from the starting point on the left and in the positive y direction from the starting point on the right. It can be seen that the y component of pointing changed in the direction of the perturbation as the starting point was varied from left to right.

In summary, the results for Experiment 3 indicate that after exposure to the perturbation subjects in the crossed x-shift and y-shift groups changed their pointing to the target in a starting point dependent manner. Although this experiment seems to indicate that subjects could adapt to multiple context-dependent perturbations of the same visual target, the results for the open x-shift condition in the direction of the perturbation were not significantly different from the controls. We hypothesized that the perturbation was not large enough to elicit significant changes in pointing and therefore repeated the open x-shift condition with a larger perturbation.

4.6 Experiment 4: Contextual Generalization of the Visuomotor Map B

Adaptation to the open x-shift condition was tested with a 10 cm perturbation and a new group of subjects. Because such a perturbation is large relative to the movement length, we also increased the length of the legs of the triangle and ran a corresponding set of control subjects.

4.6.1 Method

Subjects

16 right-handed undergraduate students participated in this study. Subjects were naive to the purpose of the experiment and were paid $7.00 for participation. All subjects had self-reported normal or corrected-to-normal vision.

Apparatus

The apparatus was identical to that used in Experiment 3.

Procedure

Subjects were randomly assigned to one of two groups: control and open x-shift. The paradigm was identical to the previous experiment except for the following changes. The perturbations for the open x-shift group were doubled in magnitude to 10 cm at the target. This required that the starting points and targets be moved to accommodate the larger perturbation. The starting locations were arranged at 7.5 cm intervals from (-17.5,10) to (27.5,10) relative to the subjects' eyes. The target was located at (5,35), the center of the grid of training points used in Experiments 1 and 2. The number of pointing repetitions was also increased from 10 to 15. Because we initially intended to measure joint angles, subjects' shoulders were fixed to the back of the

Experiment 4: Analysis of Variance							
Group	Shift	Phase		SP		Phase × SP	
		$F_{1,7}$	p	$F_{1,7}$	p	$F_{1,7}$	p
Control							
	x	< 1	ns	1.95	ns	2.38	ns
	y	< 1	ns	< 1	ns	1.70	ns
Open x-shift							
	x	2.00	ns	2.32	ns	6.40	< 0.05
	y	4.63	ns	< 1	ns	< 1	ns

Table 4.4: Summary of the two-factor within-subject ANOVAs for the two experimental groups and two directions (Dir) in Experiment 4. Non-significant effects at the $\alpha = 0.05$ level are denoted by ns.

chair with a stretch of rubber tubing. This did not interfere with subjects' ability to reach all the targets.

Analysis

The analysis was identical to that used in Experiment 3.

4.6.2 Results

Figure 4-17a shows the pattern of changes in pointing to the target as a function of the starting position for both groups. The ANOVA shows no significant changes between the pre- and post- exposure phases for the control group(Table 4.4). As in Experiment 3, the lack of interaction of phase and target in both the x and y direction indicates that there was no significant linear trend in the change in pointing as a function of starting position.

As in Experiment 3, for the open x-shift group there was a change in pointing in the y direction (Figure 4-17a). After exposure to this perturbation subjects pointed further away from the target by 1.9 cm on average. However, the ANOVA shows that the only significant effect was an interaction of phase and starting point on x pointing location. The form of this effect can be seen in Figure 4-17b. The perturbation is in the negative x direction from the starting point on the left and in the positive x direction from the starting point on the right. As the starting point is varied from left to right the adaptation in the x direction displays an increase consistent with the direction of this perturbation.

4.6.3 Discussion

For the crossed x-shift, y-shift, and open x-shift groups (in Experiment 4, with the larger perturbation), the change induced in the visuomotor map was significantly different when tested from the two starting points used in the exposure phase (2 and 6). This difference reflects a context-dependent visuomotor remapping. Although the perturbation induced from these two starting points was opposite in sign for each group, the adaptation in the visuomotor map was generally greatest at one of the two starting points while not significantly different from the controls at the other. However, when the visuomotor map was tested from the other 5 starting points the pattern of generalization reflected a smooth transition between the patterns learned at points 2 and 6. In contrast, the controls did not show any consistent pattern of change with starting position. These results suggest that not only can a context-dependent map be learned but that the map generalizes smoothly as the context is changed.

A surprising result of these experiments was the large consistent pattern of overshoot in the open x-shift groups. How can the x perturbations in these groups give rise to such a large change in pointing in the y direction? A large psychophysical (for a review see Soechting and Flanders, 1989) and neurophysiological (for a review see Georgopoulos, 1990) literature has shown a dissociation in the coding of movement direction and movement distance. Based on this dissociation, we suggest that the CNS may try to reconcile movement direction or distance when presented with conflicting information about target location. Both the distance and directions of the movement vectors from starting point 2 and starting point 6 to the target are inconsistent with the target position (Figure 4-14, solid line in open x-shift). However, the visual feedback the subject receives indicates that these vectors terminate at the same point. Therefore, to interpret the two movements as consistent with the visual input the system must resolve a conflict between the distance and direction cues given by the movements. If distance cues were to dominate, the two conflicting perturbations would be resolved into a single remapping which would be reflected in a small amount of undershoot (as the actual distance moved is less than the true distance to the target). On the other hand, if directional cues were to dominate then subjects would point to the extrapolated intersection of the two movement vectors; this would produce overshoot. Therefore, the perturbation could be resolved into a remapping with a large amount of overshoot, as was found. The results from the open x-shift experiments suggest that directional cues dominate over distance cues in this situation. However, a small amount of overshoot was seen in the crossed x-shift group which would be consistent with distance predominant over direction. It is still an open question whether these effects can be explained by differences in distance vs. direction cues.

4.7 General Discussion

In summary, we have shown spatial generalization in pointing behavior after remapping of both one and two points in the visuomotor coordinate transformation. The pattern of generalization decays away from the training points in extrinsic Cartesian coordinates. Subjects were not only able to represent two context-dependent remappings of a single point in extrinsic space but the contextual generalization showed a smooth transition as the context was varied. The implications of both patterns of generalization, as they relate to the internal representation of the visuomotor map, can be interpreted in the computational framework of function approximation.

The Function Approximation Framework

The results of these experiments can be interpreted in a computational framework by posing the problem of visuomotor learning as one of approximating the mapping between visual and motor coordinates. This mapping can be regarded as a function which transforms visual coordinates into motor coordinates. The mathematical theory of function approximation is concerned with estimating a function from samples of input-output pairs. Function approximators span a range of possible generalization patterns as measured by the behavior of the system when tested on novel inputs. Conversely, the generalization properties of a function approximator can be used to infer the internal representation of the function (Sanger, 1994). Therefore, the issues of representation and generalization are intertwined in function approximation theory.

At one extreme, a function approximator can be represented as a look-up table in which corresponding input-output pairs are stored (Atkeson, 1989; Rosenbaum et al., 1993). Thus, the visuomotor coordinate transformation could be represented as a set of pairs of visual and motor coordinates and training at one point would simply change the pairing at one location in visual space, while leaving unaltered other previously learned pairings. The substantial amount of generalization we found, however, does not support this extreme form of the look-up table model of the visuomotor map, as training at one location alters pointing to other locations.

At the other extreme of the range from local to global generalization, a coordinate transformation can be represented as a model parametrized by the physical attributes of the transformation. Thus, the motor coordinates can be represented as a function of the visual coordinates parametrized by the felt configuration of the eyes, head, and arm. In these models adaptation generally occurs through tuning of the parameters (e.g. Harris, 1965). Parametric models imply global generalization with the form of the generalization depending on the parameter that is altered. Thus, changes in the felt direction of gaze (e.g. Craske, 1967) or in felt head position (Lackner, 1973) should generalize over the whole workspace, whereas proprioceptive changes at particular loci of the arm should generalize to particular arm configurations and not others (Prablanc et al., 1975). The results from our experiments, however, do not support the

notion that the visuomotor map adapts most naturally along the previously suggested parametric lines. For example, the results obtained were qualitatively quite distinct from those predicted by changes in the felt position of the eyes or head.

Intermediate in the range of generalization ability are function approximators such as neural network models (for a reviews see Hertz et al., 1991). Neural network models fall into the general class of function approximation models that are parametrized by a large number of parameters (e.g. the weights in a neural network) that do not necessarily correspond to the physical parameters of the system. These models always predict some generalization but the extent and form of this generalization vary with the particular parameters and architecture of the model. For example, in Albus' (1975) CMAC (Cerebellar Model Articulatory Controller) model of coordinate transformations, input-output pairs are stored in a distributed fashion over a set of weights. As neighboring inputs share weights, a single point remapping will produce generalization to neighboring points in the input space with the extent determined by the overlap in weights. We will now consider a framework—regularization theory—that allows general statements to be made concerning the patterns of generalization shown by certain families of neural networks.

There is a intimate connection between the internal constraints in such network models and their resulting patterns of generalization. In general, the problem of approximating a functional mapping is severely ill-posed, since for any finite set of input-output pairs there are an infinite set of functions consistent with it. One way to resolve this problem is through the application of constraints, the topic of regularization theory (Tikhonov and Arsenin, 1977). In regularization theory, a function is approximated by minimizing a cost consisting of two terms, one characterizing the fit of the function to the input-output data and the second, known as a regularizer, denoting the preference or bias of the system with regards to a certain class of functions. Thus, for example, Bedford's linear constraint hypothesis can be interpreted as a regularizer which heavily penalizes functions with non-zero second and higher derivatives. The problem of approximating the function can now be interpreted as that of minimizing the cost; this cost plays off the match to the input-output data encoded in the first term against the intrinsic constraints of the learning system encoded in the second term.

Regularization theory can be interpreted in another way. Rather than explicitly forming the cost function and minimizing it, one can derive the form of "basis functions", or basic computational units in the function approximator, which implicitly embody the regularizer. A function formed from the data by a superposition of such basis functions minimizes the regularizer component of the cost (Poggio and Girosi, 1989). One form of basis function commonly used in neural network function approximators is the Gaussian radial basis function (Broomhead and Lowe, 1988; Moody and Darken, 1989), which can be derived by assuming a regularizer which penalizes non-smooth functions (Poggio and Girosi, 1989). Gaussian basis function networks consist of computational units with Gaussian receptive fields—units are most active

when the input is closest to the center of their receptive field, and the activity falls off in a Gaussian manner with the distance of the input to the receptive field center. Learning in such networks consists of adapting the locations and heights of the Gaussian basis units.

One consequence of representing a mapping using Gaussian basis functions is the pattern of generalization. Such basis functions will generally display a pattern of generalization which is largest at the trained point and decays in a Gaussian manner away from this point. Training at any one point only affects representational units whose Gaussian receptive fields significantly cover that point. The output of units whose receptive field centers are closest to the training point will change the most while others will change correspondingly less. We have seen in the case of the visuomotor map that the pattern of generalization is largest at the training point and decays smoothly away from it in a Gaussian manner. This suggests that, in the framework of regularization theory, the visuomotor mapping may be represented with functional basis units (e.g. radial basis functions) with large receptive fields. Furthermore, we have seen that networks with such large receptive fields embody a smoothness principle, in that training at one point changes the map gradually around that point. We therefore suggest that a spatial smoothness constraint best characterizes the patterns of generalization in the visuomotor map.

The context-dependent remapping found in Experiments 3 and 4 can be interpreted within the framework of function approximation in two ways. First, one can consider the context variable simply as another input variable in the input-output mapping that is learned. In this interpretation, the finding that a starting-position-dependent mapping of a single point in visual space could be learned suggests that the visuomotor map can be naturally parametrized by the starting position variable. The results of the generalization to other contexts further suggest that not only is the visuomotor map spatially smooth as suggested by Experiments 1 and 2, but that it is also parametrized to vary smoothly as the starting point is varied.

An alternate interpretation of the context-dependent mapping is that two separate visuomotor maps are learned and the context is used to switch between them. A suggestive computational model for how such separate modules can be learned and combined is the mixture-of-experts neural network architecture (Jacobs et al., 1991). In this model, the system starts with several modules, or experts, and learns to partition the function amongst the experts. Each expert receives a copy of the input and maps it into an output. A separate network, the gating network, weights the outputs of the experts, effectively determining which experts to rely on for each input. In the case of the visuomotor map, each expert would represent one of the mappings from visual to motor coordinates. The gating network would use the context of the movement, i.e. the starting position, to determine how to weight each expert for each context. As the context is varied, the gating network can alter the contribution of each expert network. The pattern of change in the visuomotor map as the context is varied represents the way in which the gating network generalizes in relation to the context.

Our finding of a smooth transition of the visuomotor map as the context is varied suggests, in this framework, that the gating network is gradually varying the weighting of the two experts based on the context. Again, this suggests a principle of contextual smoothness, now operating in a gating process distinct from the visuomotor maps.

Other Generalization Studies

Other than Bedford's (1989, 1993) influential work, several recent studies have addressed questions of visuomotor generalization that are relevant to our work. Imamizu et al. (1994) examined pointing under a rotation of 75° centered about a point on the table, using a setup, similar to Cunningham's (1989), in which hand movements produced cursor movements on a monitor screen. The subjects' goal was to acquire, as rapidly as possible, targets randomly presented in a circle about the initial cursor location. The authors used the duration of the ballistic portion of the movement as an indicator of learning. The results of their study indicate that learning the rotation for movements in one direction generalized to movements in the other direction. However, there are several problems with this study. First, the feedback subjects receive is on a monitor screen displaced from the finger location, therefore the cursor spot is not located at the same perceived location in space as the finger and analogies cannot be drawn to prism adaptation experiments. Second, subjects were fully informed of the nature and amount of the perturbation. Therefore, the experiment confounds perceptual and cognitive components of the task and the study consequently bears more on task learning than the representation of the visuomotor mapping.

Shadmehr and Mussa-Ivaldi (1994) studied adaptation and generalization to viscous (velocity-dependent) force fields during target directed movements. They found that exposure to such a force field in the left portion of the workspace generalized to the right portion of the workspace in joint-based, rather than Cartesian, coordinates. Our results seem to suggest that visuomotor (kinematic) learning generalizes in extrinsic Cartesian coordinates. Furthermore, there is evidence from adaptation studies on movement that Cartesian coordinates are central in planning the kinematics of arm trajectories (Wolpert et al., 1995). These results may indicate an interesting dichotomy between the representation of kinematics, in extrinsic Cartesian coordinates, and dynamics, in intrinsic joint-based coordinates.

A relevant perceptual study is that of Bülthoff and Edelman (1992), who have studied generalization in the domain of object recognition. Subjects were trained to recognize 2D views of amoeba-like objects and generalization to other poses (2D projections) of the object was assessed. They found that recognition falls off smoothly, in a Gaussian-like manner, with distance from the presented viewpoint. This finding has been taken as support for a theory of object recognition based on the superposition of basis functions, each of which represents a 2D object view (Poggio and Hurlbert, 1994).

It is interesting that our studies of visuomotor generalization show qualitatively similar effects to Bülthoff and Edelman's (1992) purely perceptual study. It may be

that the principles governing the learning and representation of mappings in the CNS transcend the particular systems involved (Poggio, 1990). For example, it is known that maps of visual and auditory space are kept in alignment in the midbrain tectum of owls (Knudsen, 1982), cats (Harris et al., 1980; Stein and Meredith, 1993), and primates (Jay and Sparks, 1984) and that prismatically imposed displacements of visual space alter the corresponding map of auditory space (Knudsen and Knudsen, 1989a). An interesting, as yet unanswered, empirical question is whether the principles of generalization that arise in the visuomotor map are also reflected in these visual and auditory maps.

The Neural Representation of the Visuomotor Map

The form and extent of generalization in the visuomotor map is intimately tied to the plasticity of the neural representations mediating this coordinate transformation. As yet the neural basis of this transformation is not fully understood, but it has become clear that the posterior parietal cortex plays a prominent role in its representation. In the posterior parietal cortex of primates, retinotopic maps have been found to be modulated by eye position in the orbit, head position relative to the body (Andersen, 1987), and most recently body orientation (Snyder et al., 1993). Computational models of this area indicate that the cell responses found neurophysiologically could be an intermediate representation in the transformation from retinotopic coordinates to extrinsic coordinates (Zipser and Andersen, 1988; Pouget and Sejnowski, 1995). Furthermore, evidence from patients with damage to this area indicates severe deficits in pointing, reaching and related spatial tasks involving visuomotor coordinate transformations (for a review see Andersen, 1987).

What do the results of our study imply for the neural representation of the visuomotor transformation? We find that changes in the mapping at one location generalize to other locations in visual space. Therefore, we expect that the neural coding of the transformation is distributed in nature and comprised of units with large functional receptive fields in visual space. Let us speculate that the visuomotor transformation is indeed computed largely by neurons in the posterior parietal cortex (PPC). Our findings are consistent with the response properties of these neurons, which have receptive fields spanning on the order of 60° of the visual field (Andersen, 1987). We further find that a single point in visual space can be remapped in two different directions depending on the starting point of the movement. As we have seen, the visual receptive fields of PPC neurons are modulated by body configuration signals such as eye position, head position, and body orientation. Furthermore, a large portion of PPC cells also display activity related to joint configuration and active arm movement (Mountcastle et al., 1975). It is plausible then, if the visuomotor map were represented in PPC, that the context of the movement, as indexed by the initial configuration and direction of arm movement, could naturally modulate it. Finally, we found that as the context was varied the visuomotor map generalized smoothly to intermediate contexts. These observed patterns of contextual generaliza-

tion may therefore reflect the effect of the arm position signals on the computation of the visuomotor mapping in the PPC.

4.8 Conclusion

We have found that the paradigms of examining spatial and contextual generalization have proved valuable in the study of the representation of the visuomotor map. From the computational perspective of function approximation theory, our experimental findings suggest that two principles, spatial and contextual smoothness, underly the representation of the visuomotor map and the constraints on its plasticity. From the neural perspective, our findings suggest that the visuomotor map is subserved by representational units with large functional receptive fields. Further study of the visuomotor map from psychophysical, neural and computational perspectives should shed light on the fascinating yet elusive phenomenon of visuomotor adaptation that has been under scrutiny for over one hundred years.

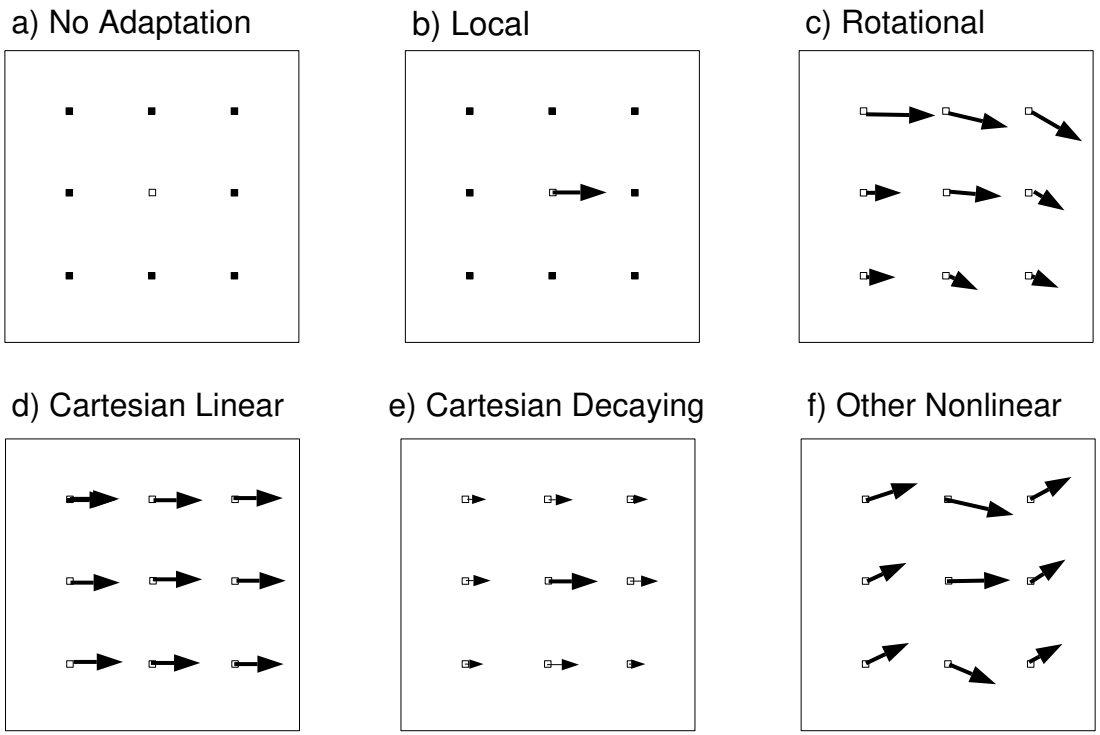

Figure 4-1: Schematic of some possible patterns of generalization to remapping of a single visual-proprioceptive pair at the central target. The central visual location is remapped to a finger position to the right of the central target. The arrows represent the change in pointing behavior after exposure to this remapping. Thus a right arrow (i.e. rightward change in pointing) at the central target represents adaptation to the perturbation at the training point whereas any change at the other eight targets is evidence of spatial generalization.

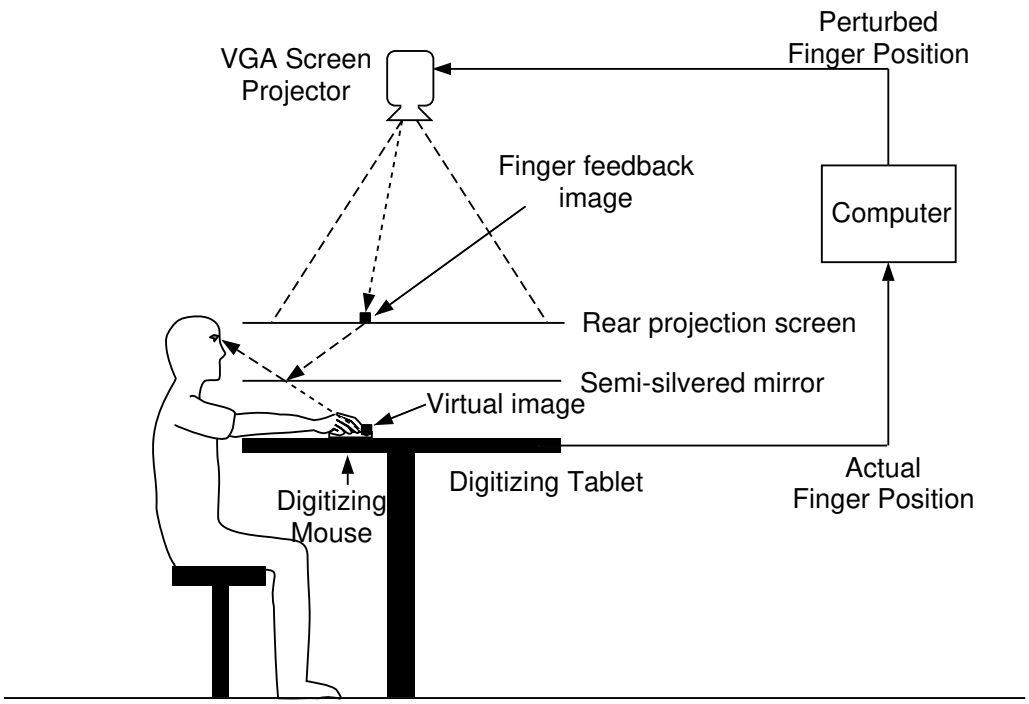

Figure 4-2: Apparatus used in all the experiments to introduce limited visuo-motor remappings. The position of the finger was captured on-line by a computer which calculated the perturbed finger position. The feedback of finger position was projected onto a screen as a cursor spot. Looking down at the mirror, the subjects saw the virtual image of the cursor spot, in the plane of the finger—the actual finger location was hidden from view. By controlling the illumination of the cursor spot the remapping could be limited to particular areas of the workspace.

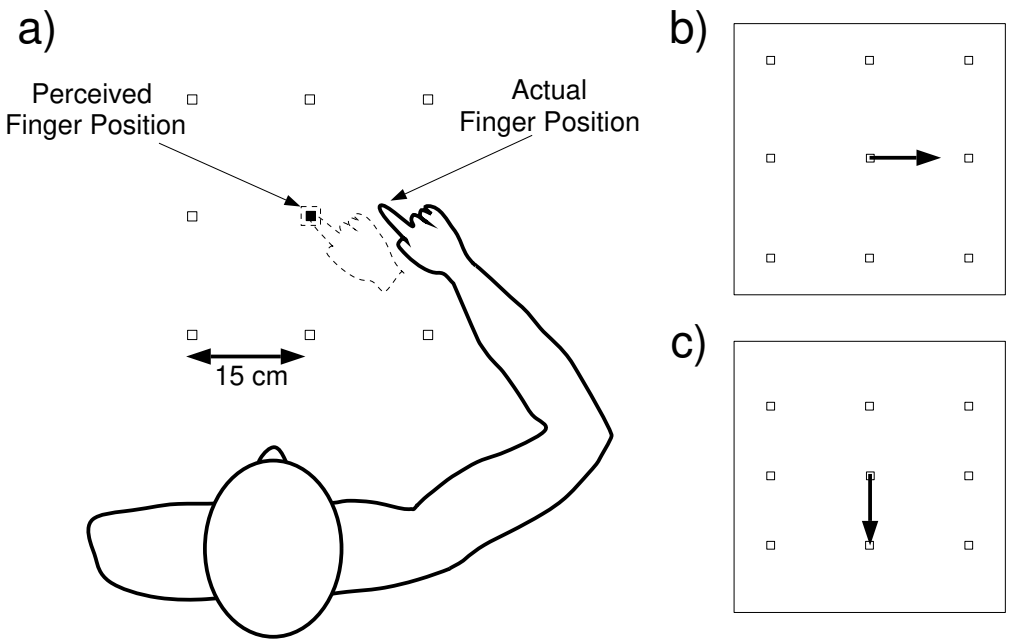

Figure 4-3: a) The position of the grid of targets is shown relative to the subject for Experiment 1. Also shown, for the x-shift condition, is the perceived and actual finger position when pointing to the central training target. The visually perceived finger position is indicated by a cursor spot which is displaced from the actual finger position. b) A schematic showing the perturbation for the x-shift group and the target numbering used to describe the results. To see the cursor spot on the central target the subjects had to place their finger at the position indicated by the tip of the arrow—a 10 cm one-point visuomotor remapping. c) A schematic similar to b) showing the perturbation for the y-shift group.

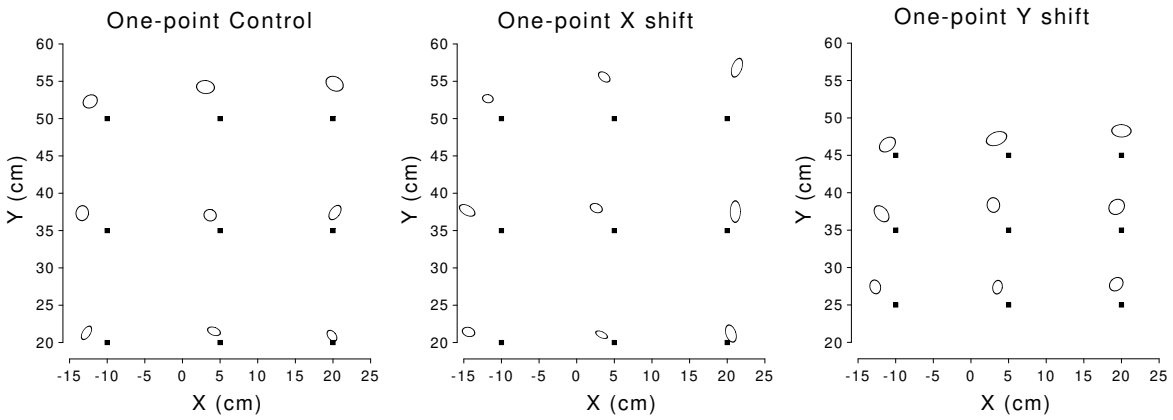

Figure 4-4: The targets (solid squares) and pre-exposure pointing locations are shown, for the three groups of Experiment 1, as 95% confidence ellipses centered around the mean.

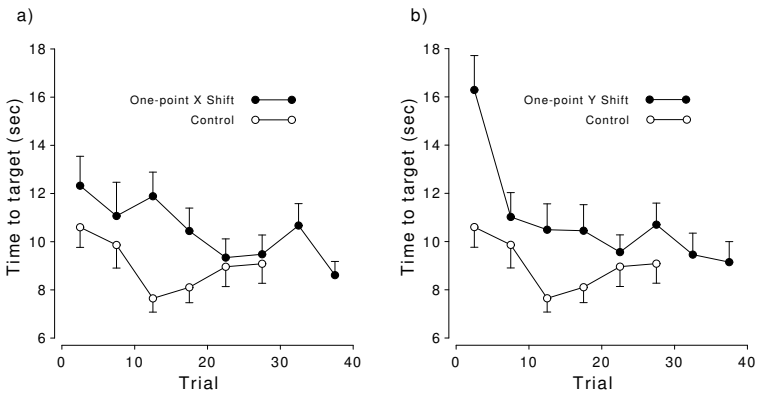

Figure 4-5: Target acquisition time as a function of trial during the exposure phase for the x-shift and y-shift groups. Also shown on each plot, for comparison, is the control group (hollow circles). For clarity the standard error bars are shown in one direction only.

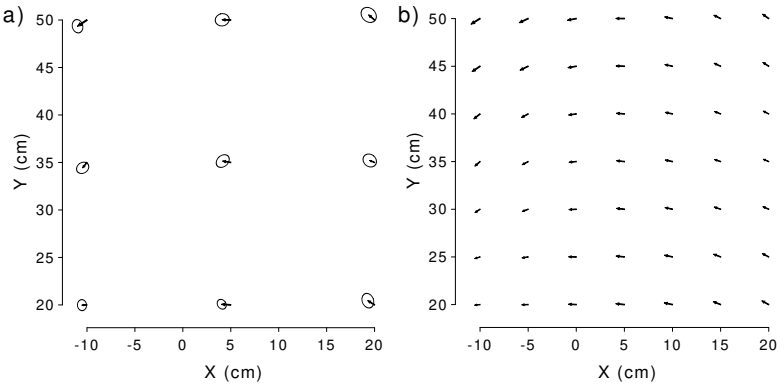

Figure 4-6: a) Average change in pointing for the control group. The arrows show the change centered on the visually presented target along with 95% confidence ellipses. b) Gaussian kernel smoothed vector field of changes.

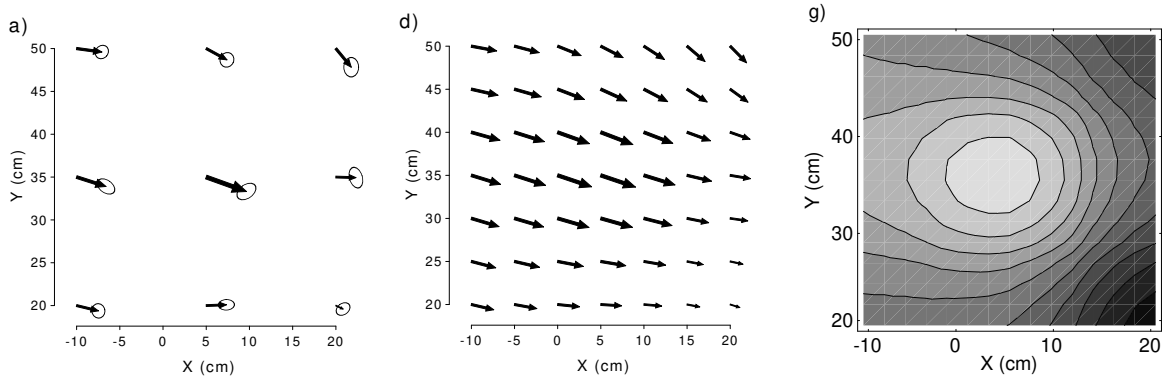

Figure 4-7: a) Average change in pointing for the x-shift group. b) Gaussian kernel smoothed vector field of changes. c) Proportion adaptation relative to the size of the perturbation. The lightest shade corresponds to 40% adaptation and the darkest shade corresponds to 11% adaptation.

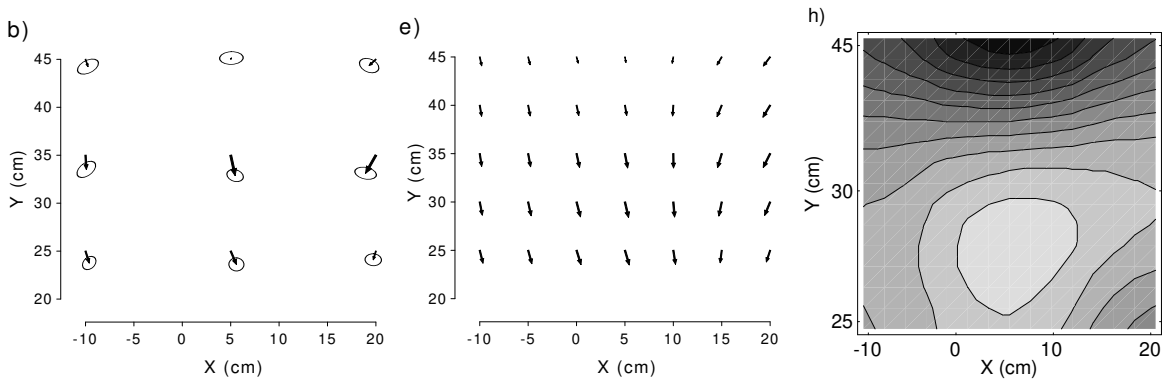

Figure 4-8: a) Average change in pointing for the y-shift group. b) Gaussian kernel smoothed vector field of changes. c) Proportion adaptation relative to the size of the perturbation. The lightest shade corresponds corresponds to 16% adaptation and the darkest shade corresponds to 6% adaptation.

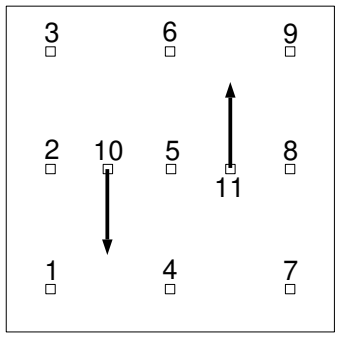

Figure 4-9: A schematic of the perturbation and the layout of the 11 targets is shown for Experiment 2. The training points and the corresponding finger positions are shown for the y-shift group. A number of global transformations are consistent with this perturbation, including a rotation and a skew.

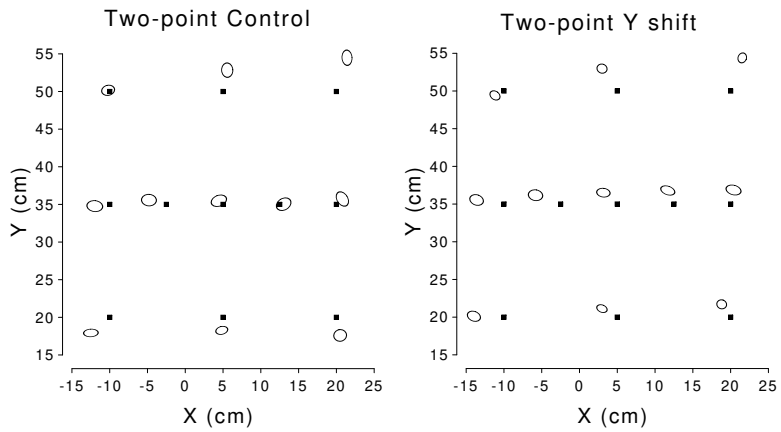

Figure 4-10: The targets (solid squares) and pre-exposure pointing locations are shown, for the two groups of Experiment 2, as 95% confidence ellipses on the mean.

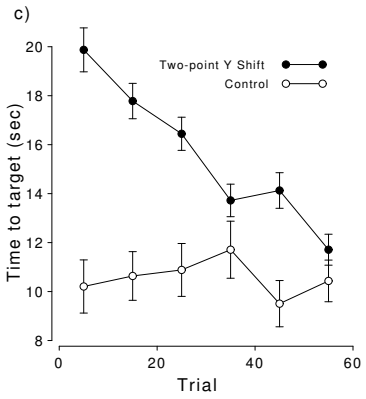

Figure 4-11: Target acquisition time as a function of trial during the exposure phase: y-shift group (solid circles) and controls (hollow circles) with one standard error bars.

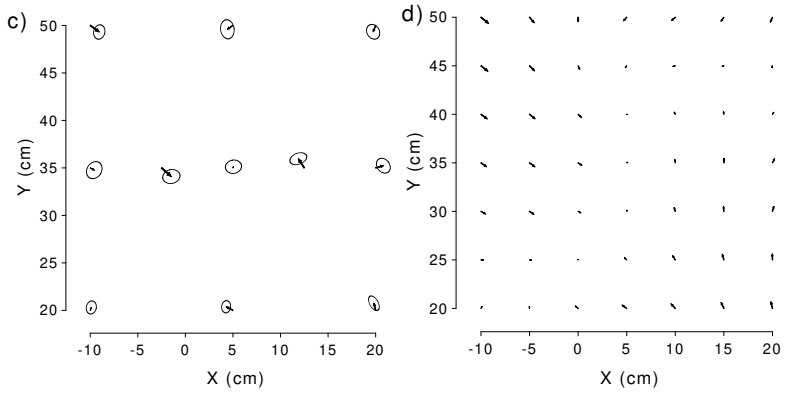

Figure 4-12: a) Average change in pointing for the Experiment 2 control group. b) Gaussian kernel smoothed vector field of changes.

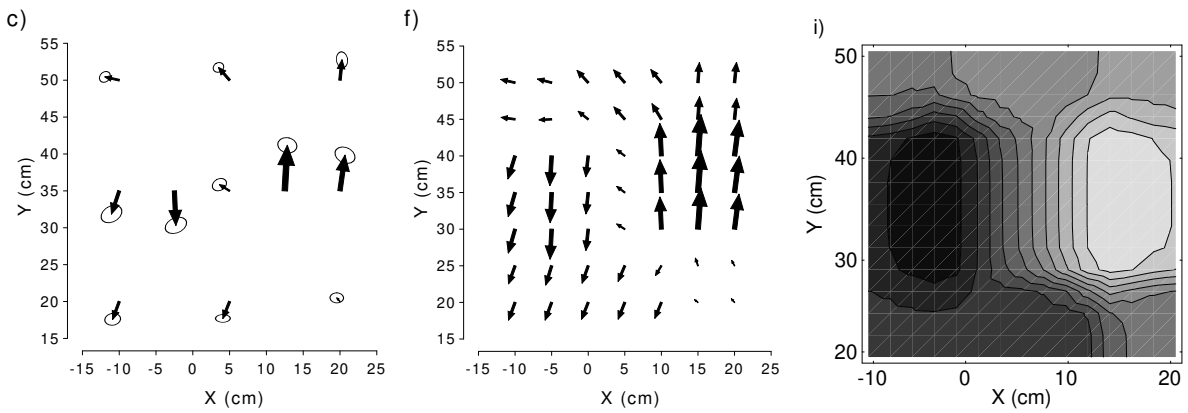

Figure 4-13: a) Average change in pointing for the y-shift group. b) Gaussian smoothed vector field of changes c) Proportion adaptation relative to the size of the perturbation. The lightest shade corresponds to 58% adaptation in the positive y direction and the darkest shade corresponds to 42% adaptation in the negative y direction.

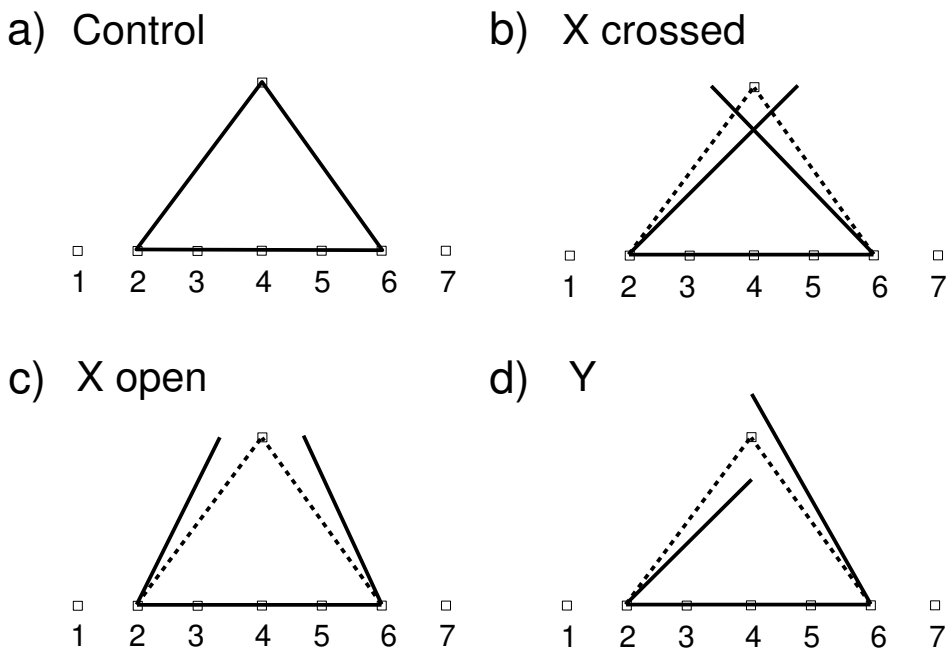

Figure 4-14: Schematic of the exposure phase of Experiment 3. The seven starting points used in the pre- and post-exposure phases are shown. During this exposure phase subjects made repeated movements to the target from starting points 2 and 6. The perturbation introduced depended on whether the movement started form start point 2 or 6. The dotted line shows the path taken by the cursor and the solid line the path taken by the finger. Note that for the three perturbation groups, although the subjects saw a triangle being traced out, the finger took a different path.

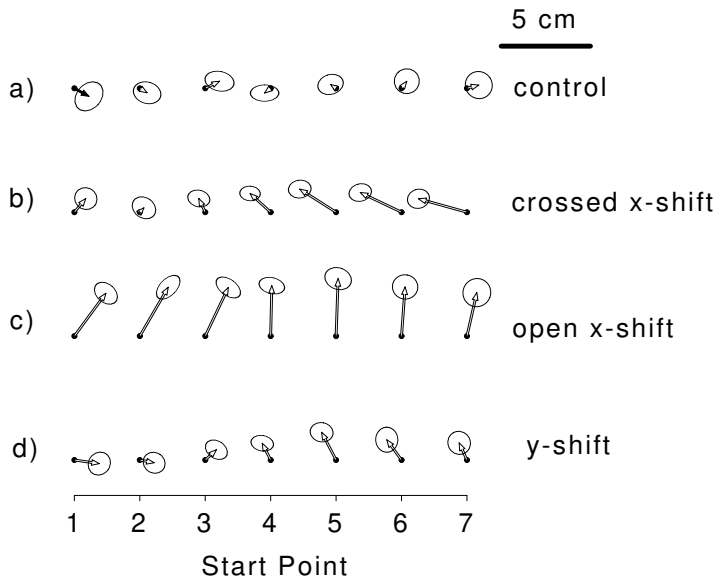

Figure 4-15: Experiment 3: Change in pointing as a function of starting position. Changes in pointing plotted as arrows with 95% confidence ellipses, arranged by starting position. All changes are calculated with respect to the target position, but are plotted as a function of starting position to clarify the starting-point dependent nature of the remapping.

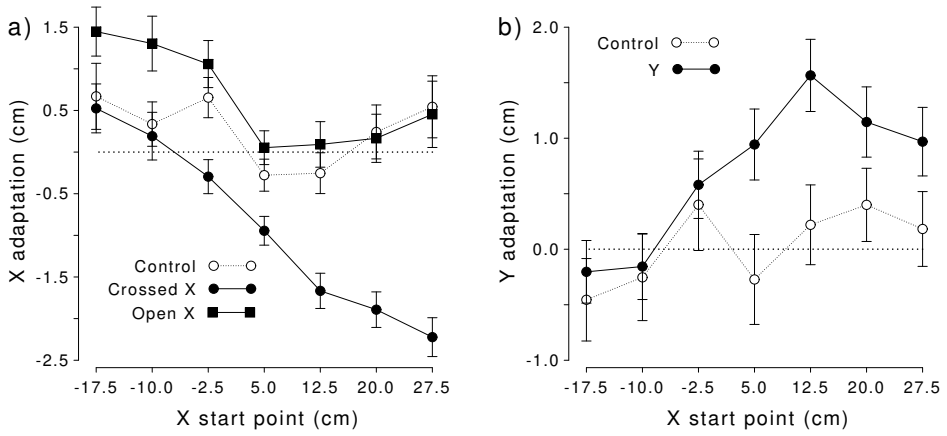

Figure 4-16: a) Adaptation in the x direction plotted as a function of starting point for the control, crossed x-shift and open x-shift groups (mean and 1 s.e.). b) Adaptation in the y direction for the control and y-shift groups.

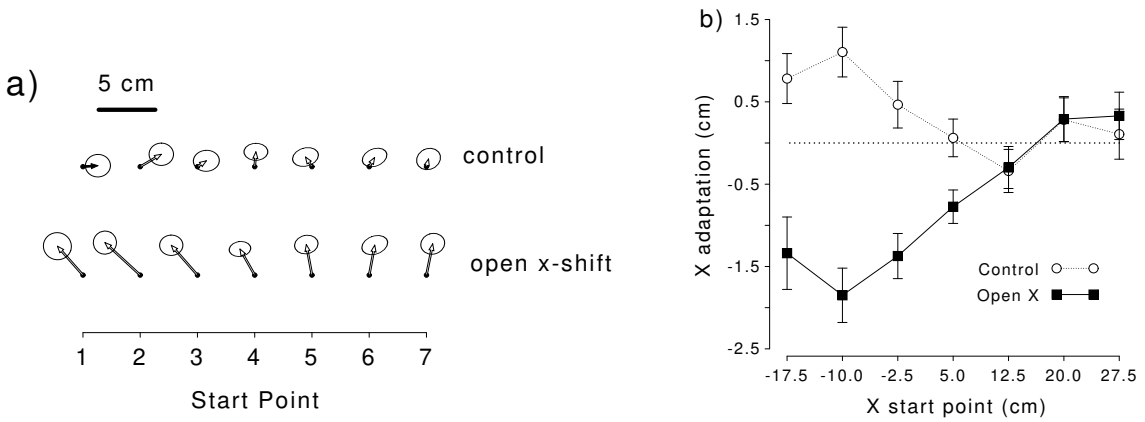

Figure 4-17: Experiment 4: Change in pointing as a function of starting position. a) Changes in pointing plotted as arrows with 95% confidence ellipses, arranged by starting position. b) Change in x direction for control and open x-shift groups.

Chapter 5

Learning Coordinate Transformations through Mutual Information

5.1 Introduction

We have reviewed evidence, throughout this thesis, that there exist multiple, mutually-aligned topographic maps in the CNS. From the perspective of multisensory integration, the existence of these maps seems computationally efficient. Once signals from multiple senses have been converted into a common coordinate system, through these aligned maps, the integration problem becomes straightforward. The coordinate transformation problem, however, remains far from trivial. In this chapter we therefore ask: How do mutually-aligned topographic maps arise in the CNS?

Two answers can be given: Aligned maps arise through innate wiring, or through experience. Evidence can be found supporting both these possibilities. For example, Aronson and Rosenbloom (1971) report that infants as young as 30 days become visibly distressed if the location of their mother's voice is displaced from the visually perceived location of the mother. Although experience in the first month of life may be crucial, this has been used to argue that visuo-auditory alignment is to some degree prewired. On the other hand, Knudsen et al. (1991) showed that complete lack of visual experience in blind-reared owls resulted in the development of maps of auditory space in the optic tectum which were stretched, upside-down, or otherwise erratic. Furthermore, Roe et al. (1990) showed that experimentally rewiring retinal inputs into the auditory pathway in ferrets induced the formation of a map of visual

space in primary auditory cortex.

In this chapter we explore, from a computational perspective, the possibility that mutually-aligned topographic maps arise purely from experience. We develop an unsupervised learning algorithm to achieve two goals: Filtering information that is common to multiple modalities, while rejecting what is not, and converting this common information into the same coordinate system. 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 goal would be to extract this commonality from other attributes, such as color, and pitch, and to generate a common map registering both visual and auditory space. What results is an algorithm which learns multiple, mutually-aligned topographic maps based purely on correlations between the inputs to the different sensory modalities.

The chapter is organized as follows. In section 5.2 we provide some background on unsupervised learning and information theory and review existing algorithms. In section 5.3 we introduce an algorithm for maximizing mutual information in topological maps and discuss its relation to previous work. In section 5.4 we demonstrate this algorithm on a small problem: learning a mapping between polar and Cartesian coordinates. We conclude the chapter with a discussion in section 5.5.

5.2 Information-theoretic Unsupervised Learning

5.2.1 Unsupervised Learning

The goal of unsupervised learning is to extract statistical structure from sensory data. From a statististical perspective, unsupervised learning is often viewed as a data modeling problem. Thus, the field of unsupervised learning has drawn heavily on statistical methods for density estimation. From the perspective of information theory, unsupervised learning is often viewed as the problem in maximizing the information content in a sensory representation. The information theoretic and statistical frameworks are closely tied; many statistical methods can be viewed as maximizing the information content in a reduced representation of the input. Models based on even the simplest information maximization rules, such as principal components analysis, have provided valuable insights into the organization of receptive fields and development of the senses (Linsker, 1988; Barlow, 1989).

The view of learning as a statistical inference problem has led to a common recipe for deriving unsupervised learning algorithms. First the sensory inputs from the environment are assumed to be generated according to some statistical model. This model—known as the *generative model*—does not need to specify details of the generation process, but it does need to constrain the space of models so as to make the problem learnable. Thus, for example, a generative model in vision may specify that contiguous patches of image have similar luminance, that corresponding patches of

the retina sense corresponding patches of space, or that the environment consists of a few translationally invariant objects. A generative model in speech may specify that signals are generated from a small set of underlying units (phonemes) constrained to transition according to a Markov chain (e.g. Juang and Rabiner, 1991).

Starting from this generative model, the learning problem consists of estimating the parameters of the model that best fit the data. This fit is generally measured by the likelihood of the data given the parameters, which can be maximized as a function of the parameters. Bayesian approaches augment this inference process by incorporating a prior distribution on the parameters, and requiring that the result of the learning process be a posterior distribution on the parameters.

A second and third framework for unsupervised learning—derived from information theory and statistical physics, respectively—can be shown to be formally equivalent to the statistical inference framework. In the information-theoretic framework, the goal of the learner is to communicate the data efficiently to a receiver, thereby producing a compact representation for the data (Cover and Thomas, 1991; Zemel, 1993; Hinton and Zemel, 1994). A cost function quantifying the efficiency of this communication process can be derived from the principle of Minimum Description Length (MDL; Rissanen, 1989). Using Shannon's coding theorem (Shannon, 1948), the MDL cost function can be shown to be equal to the posterior probability of the parameters given the data.

In the statistical physics framework, the environment and the learner are a combined system that can occupy many different *states* (for a text on statistical physics, see Parisi, 1988). The states correspond to patterns of sensory data in the environment and internal representations in the learner. An *energy* is associated with each combination of states in the environment and learner. This energy embodies the generative model: States with low energy correspond to combinations of sensory data and internal representation that have high probability, and vice-versa. The goal of learning is to find parameters of the system that minimize the energy over the entire data set. More precisely, the quantity that is minimized comprises both an energy term, indicating the fit of the parameters to the data, and an entropy term, indicating the prior probability of the parameters. Using the Boltzmann distribution, it can be shown that minimizing this quantity corresponds to maximizing the posterior probability of the parameters given the data.

Although the three frameworks are equivalent, the links to the mathematically rich disciplines of statistics, information theory, and statistical physics offer three complementary views and toolboxes for understanding learning. With this in mind, we will focus on the problem of learning coordinate transformations from the framework of information theory.

5.2.2 Information Theory

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 and Thomas, 1991). Let n be the number of possible codes for the signal, and p_j be the probability of code j. Intuitively, receiving a signal which is always predictable (e.g. $p_1 = 1$, $p_{j\neq 1} = 0$) provides no information, while receiving a signal which is maximally uncertain ($p_j = 1/n$) provides maximal information. Furthermore, the information content of the signal increases as n increases. It can be shown that the Shannon entropy

$$H = -\sum_{j=1}^{n} p_j \log p_j \tag{5.1}$$

satisfies these postulates and is unique within a constant factor (Shannon, 1948; Shannon and Weaver, 1949). For continuous signals, a limiting argument is used to convert this sum into an integral, provided that the signal cannot be observed perfectly, as this would result in infinite information. Let $H(X)$ denote the information content of random variable X, which can take on values x_j. If $p(x_i, y_j)$ is the joint probability of X taking value x_i and Y taking a value y_j, then the *joint information* for the two signals is

$$H(X,Y) = -\sum_{i,j} p(x_i, y_j) \log p(x_i, y_j).$$

Similarly, the *conditional information* of the signals is defined to be

$$H(Y|X) = -\sum_{i,j} p(x_i, y_j) \log p(y_j|x_i).$$

A transformation f of X can be considered a communication channel, receiving a signal X and outputting the transformed signal $f(X)$. The rate of transmission of information for such a channel is:

$$\begin{aligned} R &= H(X) - H(X|f(X)) \\ &= H(f(X)) - H(f(X)|X) \\ &= H(X) + H(f(X)) - H(X, f(X)) \end{aligned}$$

A transformation such as $f(X) = 0$ has a 0 rate of information transmission, whereas the identity transformation $f(X) = X$ has maximal information transmission, $R = H(X)$. The rate of information transmission between two signals is also known as *mutual information.*

5.2.3 Previous approaches

Since its inception, information theory has been an important tool for understanding the neural organization of perception. Attneave (1954, 1959) was perhaps the first to pursue the notion that visual perception could be studied from the perspective of information transmission. Barlow (1961, 1989) proposed that a specific information-theoretic criterion—minimizing redundancy—played a central role in the formation of neural representations in the perceptual system.

Serious computational modeling of the role of information theory in the formation of neural representations did not start until Linsker (1986a,b,c). His model consisted of a multilayered network of units with spatially-localized receptive fields, in which learning took place via a Hebbian rule (Hebb, 1949). (The Hebb rule states that synaptic strengths in the brain change in proportion to the correlation of the firing of pre- and post-synaptic neurons.) When a pattern with no spatial or temporal correlations was input to the lowest level of this network, the higher levels would form successively more complex representations. Among the properties of the higher levels of the model were both the on-center off-surround receptive fields and the orientation bands characteristic of primary visual cortex. Linsker's model contributed two important insights: That the principle of maximal information transmission was embodied in the simple Hebb rule, and that this principle alone operating on random inputs could account for some of the main classes of receptive fields found in primary visual cortex. Since Linsker, the idea that maximal information transmission may play a central role in early visual development has found further empirical support. Extensions of this idea, such as minimizing information loss (Plumbley and Fallside, 1988; Plumbley, 1991), have also been actively pursued.

Many researchers have pointed out that, under the simplified assumption that the inputs and outputs are Gaussian distributed, the maximum information transmission criterion embodied by the Hebb rule for linear networks performs principal components analysis (PCA; Oja, 1982, 1989, Baldi & Hornik, 1989; Sanger, 1989; Cottrell, Munro, Zipser, 1987; Plumbley, 1991). The information content of a Gaussian is proportional to the determinant of its variance; therefore maximizing the information transmission is equivalent to maximizing the variance of the output, which is exactly the goal of a linear PCA network.

Barlow's (1989) criterion of minimal redundancy, or maximal independence, has proven difficult to implement in a neurally-plausible learning rule. However, the closely related criterion of maximal de-correlation can be approximated by the combination of a Hebbian feedforward rule and an anti-Hebbian lateral inhibition rule (Barlow & Földiák, 1989). Földiák (1990) has found that this rule can form sparse representations, and, like the Linsker model, can predict some interesting receptive field properties of neurons in the early visual pathway.

All of the above algorithms are based on the idea that one of the goals of sensory processing is the preservation of information from input to output. An interesting extension of these ideas is that the goal of *multi*-sensory processing may be to maxi-

mize the mutual information between the representations in different sensory modalities (Becker, 1992; Becker and Hinton, 1992). Consider a system in which there are two modalities with inputs X and Y, and two corresponding sensory transformations $f(X)$ and $g(Y)$. The mutual information between the two transformed signals is

$$I(f(X), g(Y)) = H(f(X)) + H(g(Y)) - H(f(X), g(Y)). \tag{5.2}$$

The first two terms are the entropy of the two separate modalities. Maximizing these corresponds to maximizing the information transmission in each modality, thus recovering Linsker's criterion. The third term is the negative joint entropy of the two modalities. This term is minimized when the outputs of the two sensory transformations are maximally predictable. The two mappings, f and g, adapt so as to extract the information common to their inputs. Maximizing the mutual information between two modalities therefore corresponds to maximizing the information transmitted in each modality while maximizing predictability between their outputs.[1]

Becker & Hinton (1992) applied their algorithm to the problem of discovering disparity from random-dot stereograms (Julesz, 1971). The model consisted of two modules, each a neural network, which received inputs from corresponding patches from the two eyes. The mutual information criterion was used to derive gradient descent rules for adjusting the weights in the networks. As the only information common to corresponding patches in the two eyes was the shift due to stereo disparity, the output of the network formed a reliable representation of the stereo disparity.

This model made two important contributions to the theory of information-theoretic perceptual learning. First, it extended the simple idea of information transmission to structures with multiple modules. Second, it showed that this extension could capture interesting structure, such as stereo disparity, that is not present in inputs to any single module. Other researchers have since worked on related ideas. For example, de Sa (1994) showed how the related notion of maximizing coherence across modules could be used for classification. The basic goal in all this work is for each modality to extract the information that is common between its input and the other modality's. To this end, mutual information seems like an ideal information-theoretic criterion for multisensory integration.

[1] Just like maximizing information transmission in the linear Gaussian case reduces to the standard statistical technique of PCA, maximizing mutual information reduces to canonical correlation. The goal of canonical correlation is to find projections of two data sets that have both high variance (high information) and are maximally correlated (low joint information). See Johnson and Wichern (1992) for a text describing canonical correlation.

5.3 Topographic Mutual Information

One of the fundamental properties of coordinate transformations in the central nervous system that is not captured by models based on information transmission is topographic organization. From early visual, somatosensory, and auditory areas, to multisensory areas such as the colliculus and parietal cortex, neurons are arranged with their receptive fields forming topographic maps of the input space (Kandel et al., 1991). In multisensory areas, these topographic maps coincide, making transformation between the different coordinate frames possible. In the superior colliculus, for example, maps of visual and auditory space are aligned with the motor map for producing saccadic eye movements (Sparks and Nelson, 1987). This alignment allows the disparate representations of visual and auditory inputs to be mapped into a common motor representation.

In this chapter we propose an extension of the mutual information criterion for multisensory integration that incorporates topographic constraints. The goal is to derive an unsupervised computational model for the formation of aligned topological maps based purely on input statistics. The approach will be to first formulate the cost function which combines both mutual information maximization principles and topographic constraints, and then derive the learning algorithm which will minimize this cost function.

5.3.1 The cost function

The basic model consists of several sensory modalities, indexed by i, each of which converts an input, $\mathbf{x}_i$, into a transformed representation, $\mathbf{z}_i$ (Figure 5-1). The cost function we derive maximizes the mutual information between the $\mathbf{z}_i$, while maintaining a prespecified topographic order between and within the $\mathbf{z}_i$. If the topographic constraints between the modalities are chosen to specify a one-to-one mapping, then the transformed representations are equivalent and can be used as a common representation to convert inputs from one coordinate frame to another. Furthermore, this common representation can serve to integrate multiple modalities into a common motor pathway.

We start by expressing each of the sensory transformations as a Gaussian mixture model (McLachlan and Basford, 1988). The input to modality i is a real vector $\mathbf{x}_i$ and the transformed representation is the zero–one vector $\mathbf{z}_i$. Element $z_{ij} = 1$ if unit j is active; in a mixture model, one and only one element is allowed to be active at any one time. Given this constraint of mixture models, the total probability of the input can be written as the sum over the exhaustive and mutually-exclusive hidden representations:

$$P(\mathbf{x}_i) = \sum_{\mathbf{z}_i} P(\mathbf{x}_i|\mathbf{z}_i)P(\mathbf{z}_i). \tag{5.3}$$

For a Gaussian mixture model, the conditional probability of the input given the

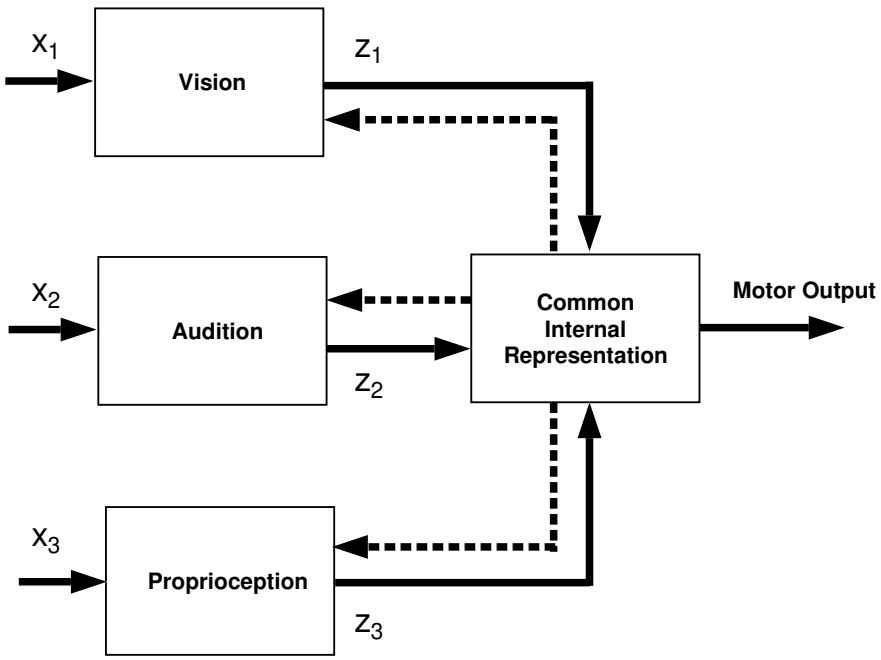

Figure 5-1: The basic model for multisensory integration, shown with some possible labels for the sensory modalities. Information from the different modalities arrives in different coordinate systems. Each modality transforms its inputs into a common coordinate system, which can then be used to integrate the inputs into a single representation. The dashed arrows represent the possible transformations, from the common representation back into the modality-specific representation, which may mediate intersensory adaptation.

hidden representation follows a Gaussian distribution:

$$P(\mathbf{x}_i|\mathbf{z}_i) \propto \exp\{-\frac{1}{2}(\mathbf{x}_i - W_i\mathbf{z}_i)^T V_i^{-1}(\mathbf{x}_i - W_i\mathbf{z}_i)\} \tag{5.4}$$

where W_i is the matrix whose rows are the means of the Gaussians and V_i is the covariance matrix common to all the Gaussians. The overall cost function will be defined in terms of the negative log likelihood, or *energy*, of the model. Taking the negative of the log of (5.4) we obtain the first term in the cost function:

$$E_i^{\text{mix}} = \frac{1}{2}(\mathbf{x}_i - W_i\mathbf{z}_i)^T V_i^{-1}(\mathbf{x}_i - W_i\mathbf{z}_i). \tag{5.5}$$

The Gaussian mixture model defined by (5.3) and (5.4) is a statistical formalization of a layer of units with Gaussian receptive fields in which both the centers and sizes of the receptive fields adapt competitively to capture the distribution of the data (Nowlan, 1991). As a model of competitive learning or clustering, it is closely related to von der Malsburg's (1973) model of self-organization, the Neocognitron (Fukushima, 1975; Fukushima, 1980), Kohonen's feature maps (1982, 1989), Adaptive Resonance Theory (ART; Grossberg, 1987; Carpenter & Grossberg, 1988) and Rumelhart and Zipser's (1985) competitive learning model. The Gaussian mixture model, however, is firmly grounded in statistics, and can, therefore, be naturally expressed in the equivalent language of information theory. In information processing terms, a mixture model converts a continuous signal into a discrete code whose information content is related to the posterior probability of the codes through the free energy (Hinton and Zemel, 1994).

Under cost function (5.5) the units in each mixture model are not arranged topographically. We can induce a topographic order by adding the following term to the cost function:

$$E_i^{\text{top}} = \frac{1}{2}\sum_{jk}[\Psi_i]_{jk}(W_{ij} - W_{ik})^T V_i^{-1}(W_{ij} - W_{ik}). \tag{5.6}$$

The matrix Ψ_i encodes the topography of the map; its elements $[\Psi_i]_{jk}$ are inversely related to the distance between unit j and k. This term penalizes units nearby on the topographic map for having the centers of their receptive fields far away in input space. By choice of the values in Ψ_i, a 1-D lattice, 2-D lattice, circular or other topologies[2] can be induced. We will restrict our attention to 2-D lattices as they are most representative of the layered maps found in many areas of the CNS.

The topographic term (5.6) is identical to the cost minimized by elastic networks

[2]Whereas the *topography* defines distances between neighbors, the *topology* defines only the neighborhood relations and is invariant to arbitrary stretching or scaling.

(Durbin and Willshaw, 1987; Yuille, 1990) and closely related to the cost implicitly minimized by Kohonen's feature maps (1982, 1989). It also has an interesting information-theoretic interpretation. The matrix Ψ_i encodes the confusability of vector $\mathbf{z}_i$, i.e. the probability of transmitting an incorrect code k when the true code is j. When Ψ_i is symmetric, it defines a distance metric over the space of codes. This confusability matrix therefore implicitly defines a topology in a system whose goal is maximizing information transmission (Luttrell, 1989, 1994; G.E. Hinton, personal communication).

The topographic term (5.6) introduces constraints within each sensory modality. We introduce a topographic constraint *between* two modalities via the term:

$$E^{\text{align}} = \frac{1}{2}\mathbf{z}_i^T \Psi_{i\ell} \mathbf{z}_\ell. \tag{5.7}$$

This term acts to align pairs of modalities, with the matrix $\Psi_{i\ell}$ playing a role very similar to Ψ_i. Intuitively, the elements of $\Psi_{i\ell}$ encode the confusability of codes across the two modalities. If $\Psi_{i\ell}$ is the negative identity matrix, for example, codes where $\mathbf{z}_i = j$ when $\mathbf{z}_\ell = j$, are favored relative to all other codes.

Finally, we introduce into the cost function the mutual information between the outputs of two modalities:

$$E^{\text{mut}} = -\mathbf{z}_i^T \log \boldsymbol{\pi}_i - \mathbf{z}_\ell^T \log \boldsymbol{\pi}_\ell + \mathbf{z}_i^T \log \Pi_{i\ell} \mathbf{z}_\ell. \tag{5.8}$$

The first two terms are the information transmitted within each modality with respect to the prior probabilities $\boldsymbol{\pi}_i$ and $\boldsymbol{\pi}_\ell$; the last term is the negative joint information with respect to the prior $\Pi_{i\ell}$.[3]

To summarize, the total energy is composed of

$$\begin{aligned} E(\mathbf{x}_1, \mathbf{x}_2, \mathbf{z}_1, \mathbf{z}_2) &= E_1^{\text{mix}} + E_2^{\text{mix}} + \alpha E_1^{\text{top}} + \alpha E_2^{\text{top}} + \beta E^{\text{align}} + E^{\text{mut}} \qquad (5.10) \\ &= \frac{1}{2}(\mathbf{x}_1 - W_1\mathbf{z}_1)^T V_1^{-1}(\mathbf{x}_1 - W_1\mathbf{z}_1) + \frac{1}{2}(\mathbf{x}_2 - W_2\mathbf{z}_2)^T V_2^{-1}(\mathbf{x}_2 - W_2\mathbf{z}_2) \\ &+ \frac{1}{2}\alpha \sum_{ij} [\Psi_1]_{ij} (W_{1i} - W_{1j})^T V_1^{-1} (W_{1i} - W_{1j}) \end{aligned}$$

[3] This can be generalized to more than two modalities by using the mutual information of multiple signals X_i, $i = 1, \ldots, n$, which is defined as

$$I(X_1, \ldots, X_n) = \sum_i H(X_i) - \sum_{ij} H(X_i, X_j) + \ldots + (-1)^{n+1} H(X_1, \ldots, X_n). \tag{5.9}$$

Such a generalization is often impractical as it requires modeling the n^{th} order statistics of the $\mathbf{z}_i$. In practice, maximizing the pairwise mutual informations may be a suitable alternative.

$$+ \quad \frac{1}{2}\alpha \sum_{ij} [\Psi_2]_{ij} (W_{2i} - W_{2j})^T V_2^{-1} (W_{2i} - W_{2j}) + \beta \frac{1}{2} \mathbf{z}_1^T \Psi_{12} \mathbf{z}_2$$
$$- \quad \mathbf{z}_1^T \log \boldsymbol{\pi}_1 - \mathbf{z}_2^T \log \boldsymbol{\pi}_2 + \mathbf{z}_1^T \log \Pi_{12} \mathbf{z}_2$$

The first two terms capture the mapping from inputs to transformed representations; the next three terms capture topographic constraints within and between the maps; and the last three terms are the mutual information; α and β set the relative importance of these terms.

5.3.2 The learning algorithm

The goal of the learning algorithm is to minimize the cost (5.10). The traditional approach to learning is based on gradient descent: The cost is minimized by taking its gradient with respect to the parameters and changing the parameters by a small amount in the direction of this gradient at each time step. Previous approaches to mutual information have found this method prohibitively slow, even for accelerated methods such as conjugate gradient descent (G.E. Hinton and P. Dayan, personal communication). We derive an alternative learning algorithm for this architecture based on the EM algorithm (Dempster et al., 1977).

The EM algorithm relies on the natural separation of variables into observables, the inputs $\mathbf{x}_i$, and hidden variables, the representations $\mathbf{z}_i$. If we assume that the model parameters are correct, the hidden variables can often be easily estimated. Conversely, if we assume that the values of the hidden variables are known, then the problem of estimating the model parameters often becomes trivial. The EM algorithm iterates between assuming correct parameters and computing the expectations of the hidden variables (E step), and using these expectations to find new parameters that maximize the expected likelihood (M step).

The energy defined in equation (5.10) can be related to a probability model via the Boltzmann equation,

$$P(\mathbf{x}_1, \mathbf{x}_2, \mathbf{z}_1, \mathbf{z}_2) = \frac{1}{Z} \exp\{-E(\mathbf{x}_1, \mathbf{x}_2, \mathbf{z}_1, \mathbf{z}_2)\}, \tag{5.11}$$

where Z is the normalization constant, also known as the *partition function*. The EM algorithm maximizes the expected log likelihood of the parameters ϕ',

$$Q(\phi'|\phi) = \langle -E(\mathbf{x}_1, \mathbf{x}_2, \mathbf{z}_1, \mathbf{z}_2) - \log Z \rangle_c, \tag{5.12}$$

where ϕ denotes the current parameters, $\phi = \{W_1, W_2, V_1, V_2, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \Pi\}$, and $\langle \cdot \rangle_c$ denotes expectation given the inputs and ϕ. The parameters, $\Psi_1, \Psi_2, \Psi_{12}$, are fixed by

the prespecified topographic structure of the network. If the Gaussians are normalized to account for the variances, and the priors $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \Pi$ are normalized to satisfy probability constraints, then a global computation of the partition function is unnecessary as it does not depend on the parameters. Therefore, unlike the Boltzmann machine (Ackley et al., 1985), an unclamped phase of learning is unnecessary. A discussion of fast single-phase learning in this and associated models with constant partition functions can be found in (Neal, 1992) and (Ghahramani, 1995).

The E step of the algorithm computes the expected log likelihood of the parameters. Like in the Boltzmann machine, this amounts to calculating the first and second order statistics, $\langle \mathbf{z}_1 \rangle_c$, $\langle \mathbf{z}_1 \rangle_c$, and $\langle \mathbf{z}_1 \mathbf{z}_2^T \rangle_c$. For the examples in this chapter, these expectations were evaluated exactly. For m modalities, each with a having k-element hidden vector, the exact E step is an $\mathcal{O}(k^m)$ computation. More efficient approximations can be obtained using Gibbs sampling (Geman and Geman, 1984), and mean field theory (Parisi, 1988). The M step of the algorithm uses the expectations calculated in the E step to estimate a new set of parameters. A detailed derivation of the EM algorithm for this architecture is given in Appendix A of this chapter.

5.4 Experiment: Polar and Cartesian Coordinates

The algorithm for maximizing topographic mutual information was tested on a small coordinate transformation problem. The problem consisted of extracting a common representation from two input modalities: One coding stimulus locations in Cartesian coordinates, (x, y), and the other in polar coordinates, (r, θ), where

$$\begin{aligned} r &= \sqrt{x^2 + y^2} \\ \theta &= \tan^{-1}(y/x). \end{aligned}$$

Polar stimulus coordinates were input into one network, and the corresponding Cartesian stimulus coordinates were simultaneously input into a second network. Each network consisted of 25 Gaussian units arranged in a 5×5 map, and there were a total of 100 polar-Cartesian input pairs. The networks were trained both with the topographic mutual information cost function (equation (5.10) with $\alpha = 0.1, \beta = 16$), and with a non-topographic control ($\alpha = \beta = 0$). The two parameters of the topographic cost function, α and β, were set by trial and error, and reflect the importance of preserving between- and within-map topographic order relative to maximizing mutual information.

Both algorithms rapidly formed representations of the two modalities with high mutual information (Figure 5-2). Whereas the topographic constraints aided the generation of high mutual information representations early in learning, the non-topographic algorithm rapidly caught up and in some simulations surpassed the to-

pographic cost function in terms of mutual information (not shown).

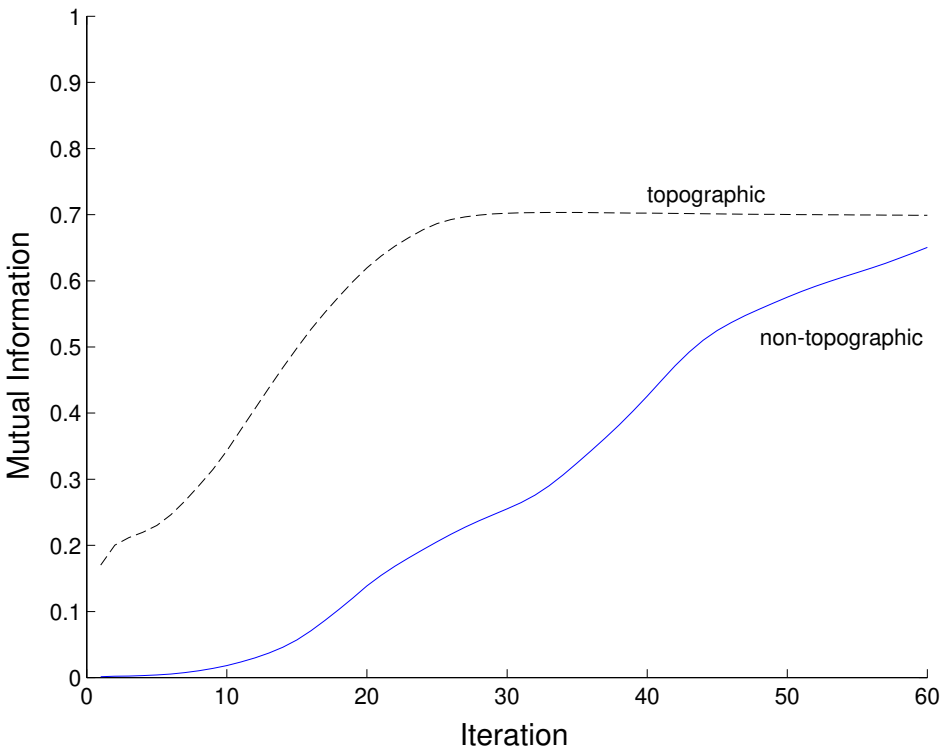

Figure 5-2: Learning curves for non-topographic (solid) and topographic (dashed) mutual information algorithms. The mutual information as a fraction of maximum possible information ($\log_2 25 = 4.64$ bits) is plotted as a function of iterations of the EM algorithm. Each iteration is one pass through the data set.

At the end of learning, the non-topographic algorithm showed all the signatures of high mutual information (Figure 5-3 bottom row). Both the polar and Cartesian representations had high information content: The receptive field centers had spread to capture the input distribution. Moreover, activity in the polar map was highly predictable from activity in the Cartesian map and vice-versa, as evidenced by the sparse joint probability matrix. However, there was no sign of topographic order, either within each of the two maps, or between them.

On the other hand, the topographic algorithm showed both high mutual information and topographic order at the end of learning (Figure 5-4 bottom row). The representations that emerged were map-like and orderly within and between the two modalities.

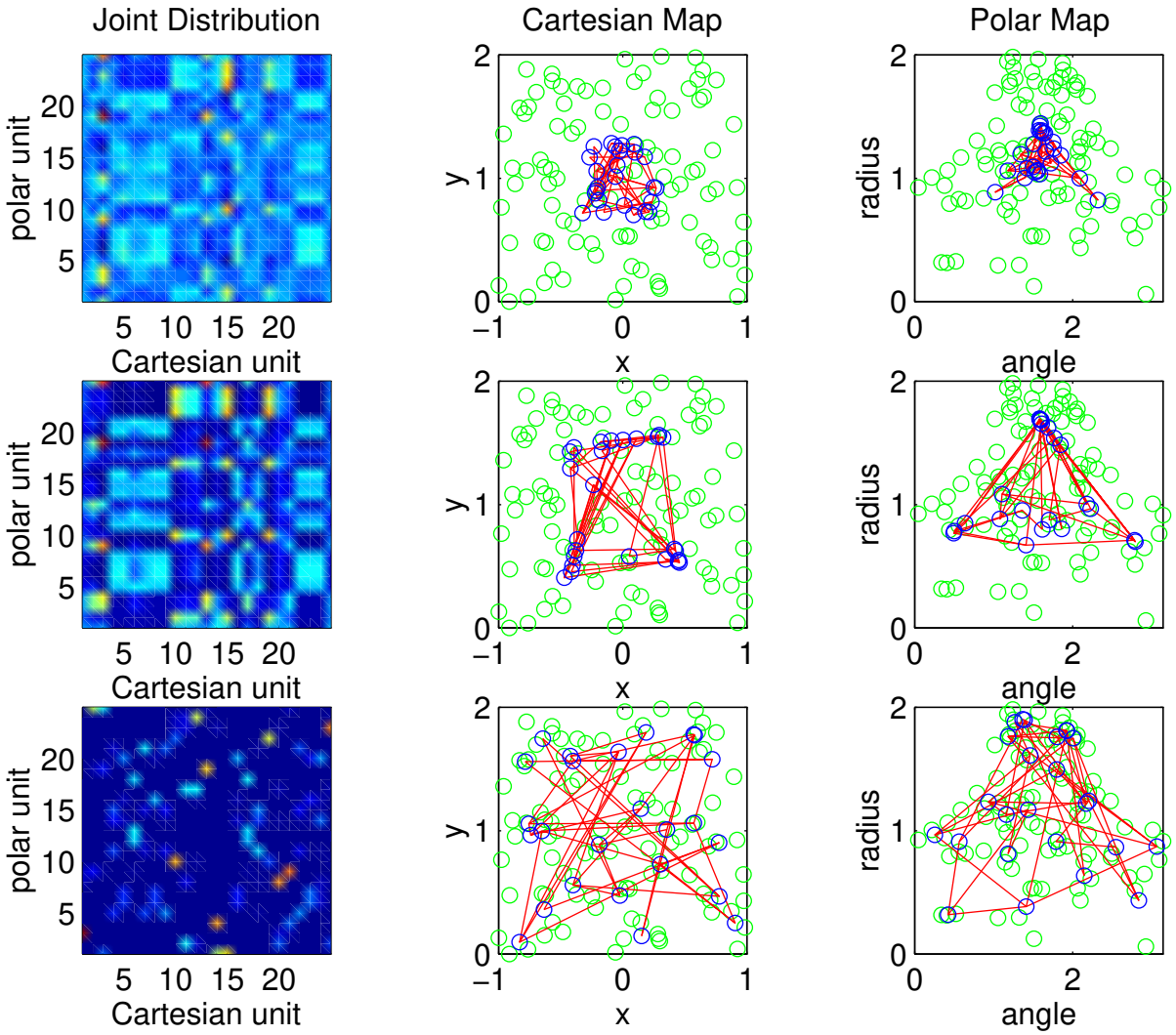

Figure 5-3: Learning in the non-topographic mutual information algorithm. Each row is a snap-shot at a different stage of learning: early (iteration 5; top), middle (iteration 20; middle), and late (iteration 60; bottom). The first column displays the joint probability distribution for the activity of corresponding pairs of units in the two maps—with lighter shades indicating high probability of joint activity and darker shades indicating low probability. The second and third columns display the arrangement of receptive fields in the Cartesian and polar maps, respectively. The dark circles are receptive field centers; the light circles are the data points (sensory inputs); the lines indicate neighborhood relations between receptive fields.

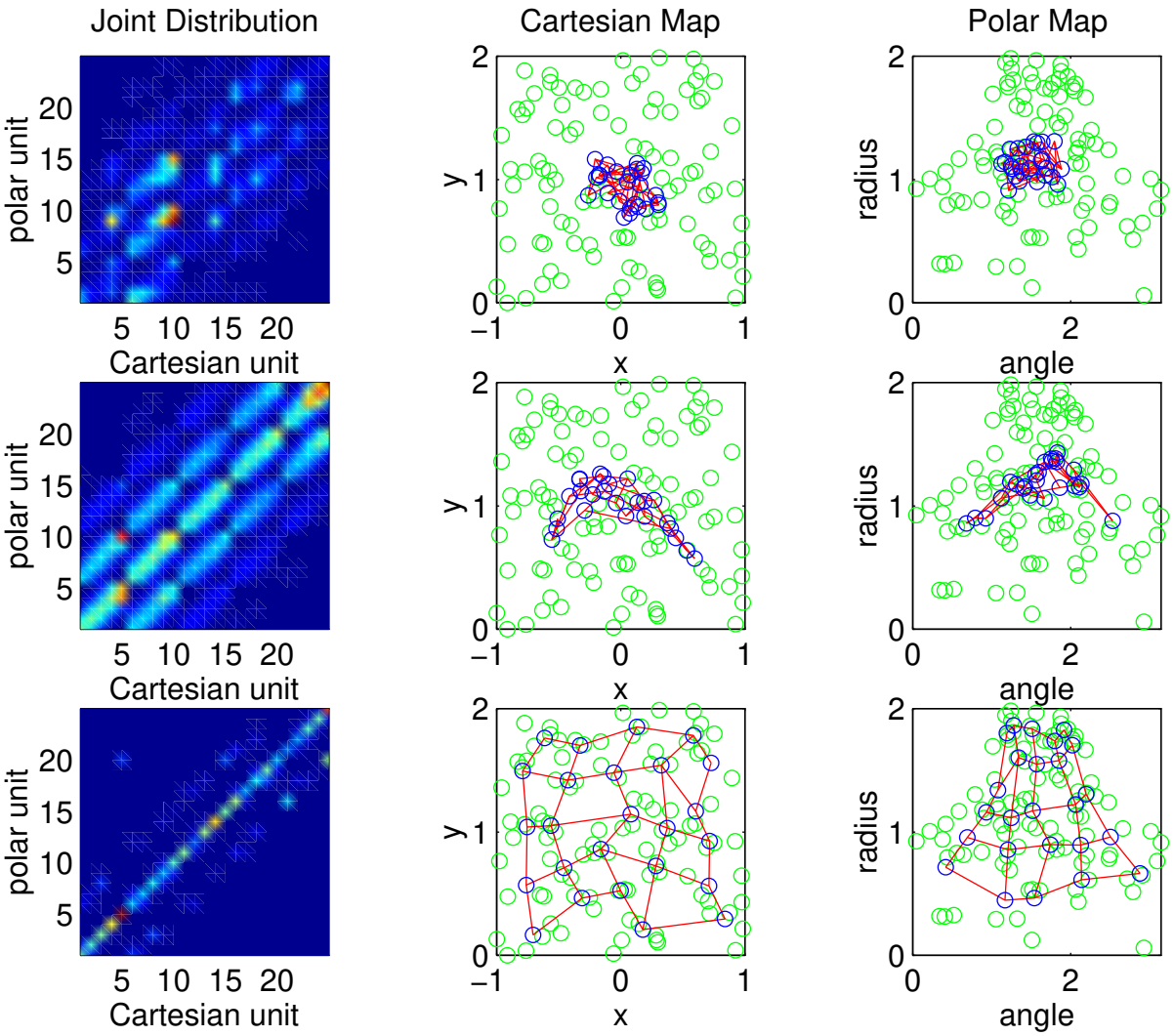

Figure 5-4: Learning in the topographic mutual information algorithm. Each row is a snap-shot at a different stage of learning: early (iteration 1; top), middle (iteration 5; middle), and late (iteration 60; bottom). The plots are arranged as in the previous figure.

5.5 Discussion

In this chapter we have derived an information-theoretic criterion for learning common representations of multiple modalities. The criterion is based on maximizing the mutual information between the representations in each sensory modality. We augment this criterion by imposing topographic structure within and between the sensory modalities. The representations derived from this new criterion can be used both to transform between representations in the different modalities, and as a common representation subserving motor output.

The Polar–Cartesian problem illustrated that although mutual information can be maximized without any topographic constraints, the representations derived do not reflect the structure of the input space. Highly structured representations can be obtained by adding the appropriate terms. In future simulations we plan to explore the problem of transforming visual and auditory inputs into a common spatial map.

It should be noted that the topographic structure imposed between and within sensory modalities may be different. For example, a two-dimensional map (e.g. interaural time difference (ITD) and interaural intensity difference (IID)) may be mapped onto a one-dimensional continuum (location in azimuth) in another modality. More esoteric structures, such as the helical structure of perceived tones revealed by similarity judgements (Shepard, 1982), may also be induced through combinations of circular and linear topologies.

The algorithm can be extended to more than two modalities by use of the n-channel generalization of mutual information (5.9). Generalizing to n modalities requires both the estimation of n^{th} order statistics of the hidden representation in the E step, and the corresponding n^{th} order parameters in the M step. Both computation time and overfitting problems render such a generalization infeasible. A natural tractable approximation can be obtained by truncating (5.9) to include only the first and second order joint entropy terms, thereby reducing the mutual information cost function to just pairwise terms. This approach has not been tried.

It may seem that the algorithm derived in Appendix A is so biologically farfetched it has little relevance to the natural learning and development of coordinate transformations that takes place in the CNS. The data are all processed in batch; the algorithm relies on matrix inversion and relatively complex linear algebra; operations are non-local, etc. However, inspection of the cost function shows that it is composed primarily of quadratic terms relating the different representations. These can be exactly implemented through Gaussian receptive fields and Gaussian connectivity patterns among hidden units. A stochastic gradient descent algorithm applied to this architecture results in a learning rule which moves these receptive fields in the direction of inputs, receptive fields of nearby units in the same map, and receptive fields of corresponding units in the other sensory map. Therefore, although the present implementation of the algorithm is biologically implausible, it can probably be implemented using learning rules based on receptive field plasticity and changes in connectivity. Plasticity of receptive fields and changes in connectivity are not only

neurally plausible but closely linked to many types of experience-dependent plasticity, such as that found in the visual (Wiesel and Hubel, 1963), somatosensory (Merzenich et al., 1983), and motor systems (Sanes et al., 1990; Donaghue et al., 1990). Using these mechanisms, a model based on maximizing mutual information in topographic maps may capture some basic properties of experience-dependent development of mutually-aligned maps.

In conclusion, an unsupervised algorithm for learning coordinate transformations has been derived from information-theoretic terms. This algorithm combines the principles of maximizing mutual information and preserving topographic structure. Many questions for further research arise from this framework. We will briefly outline three of these questions.

First, the algorithm provides a natural model for the psychophysics of multisensory integration and adaptation (cf. Chapters 2 and 4). For example, the model can predict, based on information-theoretic principles, the effect of varying amounts of noise in the input modalities. These quantitative predictions apply to all three conditions studied in Chapter 2: Fusion of unperturbed signals, adaptation to added bias, and adaptation to added variance. Comparison of the empirical results in Chapter 2 with the predictions made by this model therefore seems the next natural step in this research project.

Second, the model roughly captures some of the essential features of multisensory areas such as the superior colliculus: topology, alignment, and adaptability. It may be fruitful to make this analogy more explicit. For example, the receptive fields and connectivity within these multisensory maps could be used to predict the structure that would emerge from the topographic mutual information criterion. Furthermore, it would be fruitful to consider whether a local Hebbian-like learning rule, coupled with the physiologically observed connectivity, could implement the topographic mutual information cost function.

Third, from a purely computational standpoint, it is important to derive more tractable approximations to the EM algorithm derived in this chapter. The main problem is that in large networks the exact E step is computationally expensive. It should be fairly straightforward to derive both a stochastic version of the algorithm based on Gibbs sampling, and a deterministic variant based on mean field theory.

Appendix A: Derivation of the EM algorithm

The E step of the algorithm computes the expected log likelihood $Q(\phi'|\phi)$ of the parameters. For this model, dropping the terms that do not play into the maximization and noting that $\langle \mathbf{z}_i^T \mathbf{z}_i \rangle_c = 1$,

$$Q(\phi'|\phi) = -\mathbf{x}_1 V_1^{-1} W_1 \langle \mathbf{z}_1 \rangle_c - \mathbf{x}_2 V_2^{-1} W_2 \langle \mathbf{z}_2 \rangle_c + \frac{1}{2} W_1^T V_1^{-1} W_1 + \frac{1}{2} W_2^T V_2^{-1} W_2$$

$$
\begin{aligned}
&+ \quad \frac{1}{2}\alpha \sum_{ij} [\Psi_1]_{ij} (W_{1i} - W_{1j})^T V_1^{-1} (W_{1i} - W_{1j}) \\
&+ \quad \frac{1}{2}\alpha \sum_{ij} [\Psi_2]_{ij} (W_{2i} - W_{2j})^T V_2^{-1} (W_{2i} - W_{2j}) \\
&+ \quad \frac{1}{2}\beta \mathrm{tr}[\Psi_{12}^T \langle \mathbf{z}_1 \mathbf{z}_2^T \rangle_c] - \langle \mathbf{z}_1 \rangle_c^T \log \boldsymbol{\pi}_1 - \langle \mathbf{z}_2 \rangle_c^T \log \boldsymbol{\pi}_2 + \mathrm{tr}[\log \Pi_{12}^T \langle \mathbf{z}_1 \mathbf{z}_2^T \rangle_c].
\end{aligned}
$$

The E step therefore relies on calculating the first and second order statistics, $\langle \mathbf{z}_1 \rangle_c$, $\langle \mathbf{z}_1 \rangle_c$, and $\langle \mathbf{z}_1 \mathbf{z}_2^T \rangle_c$.

The M step maximizes the expected log likelihood with respect to the parameters. Assume the data set consists of N patterns $\{\mathbf{x}_{in}\}_{n=1}^N$. Setting the derivatives of Q with respect to the mean vectors to zero, we obtain a linear system of equations

$$
\frac{\partial Q}{\partial W_{ij}} = V_i^{-1} (\sum_n \mathbf{x}_{in} \langle \mathbf{z}_{ijn} \rangle_c^T - W_{ij} \sum_n \langle \mathbf{z}_{ijn} \rangle_c^T) - 2\alpha N \sum_k [\Psi_i]_{jk} V_i^{-1} (W_{ij} - W_{ik}) = 0,
$$

for W_i. The solution to this gives

$$
\widehat{W_i} = \Lambda^{-1} \sum_n \mathbf{x}_{in} \langle \mathbf{z}_{in} \rangle_c^T,
$$

where $\Lambda_{jk} = \delta_{jk} (\sum_n \langle \mathbf{z}_{ijn} \rangle_c + 2\alpha N \sum_l [\Psi_i]_{jl}) - 2\alpha N [\Psi_i]_{jk}$, and δ_{jk} is the Kronecker delta.

To estimate V_i, we solve $\dfrac{\partial Q}{\partial V_i} = 0$ obtaining the linear equation

$$
\widehat{V_i} = \frac{1}{\lambda} \sum_{nj} \langle \mathbf{z}_{ijn} \rangle_c (\mathbf{x}_{in} - W_{ij})(\mathbf{x}_{in} - W_{ij})^T + \frac{2\alpha N}{\lambda} \sum_{jk} [\Psi_i]_{jk} (W_i j - W_i k)(W_i j - W_i k)^T
$$

where $\lambda = N(1 + 2\alpha \sum_{jk} [\Psi_i]_{jk})$.

To estimate $\boldsymbol{\pi}_i$, we solve $\dfrac{\partial Q}{\partial \boldsymbol{\pi}_i} = 0$ obtaining

$$
\widehat{\boldsymbol{\pi}_i} = \sum_n \langle \mathbf{z}_i n \rangle_c.
$$

In the simulations in this chapter we kept $\boldsymbol{\pi}$ fixed to a maximum entropy distribution, $\boldsymbol{\pi}_{ij} = 1/k$.

Estimating Π_{12} is more difficult. Unlike all the other parameters, for which the

expected log likelihood yields a single global maximum, this parameter plays into the cost with a negative sign; setting $\frac{\partial Q}{\partial \Pi_{12}} = 0$ yields the global *minimum*. In fact, the maxima lie at the boundaries of a simplex defined by the constraints that $\sum_j [Pi_{12}]_{ij} = \boldsymbol{\pi}_{1i}$ and $\sum_i [Pi_{12}]_{ij} = \boldsymbol{\pi}_{2j}$. This suggests using linear programming (Press et al., 1988) to solve the M step. The simpler method employed in this chapter is to use a partial (gradient) M step for this parameter,

$$\Delta[\Pi_{12}]_{jk} = -\eta \frac{\langle \mathbf{z}_{1j} \mathbf{z}_{2k} \rangle_c}{[\Pi_{12}]_{jk}}.$$

Appendix B: Matlab Code for Topographic Mutual Information

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Maximizing Mutual Information Between Two Gaussian Mixtures        %
%                                                                    %
% Zoubin Ghahramani 6/20/95.                                         %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

disp('loading training data');
load data -ascii;
inputs=data;
sim=1;
K=5;                    % size of mesh
M=K*K;  % number of Gaussians
D=length(inputs(1,:))/2;% dimensionality of the input
N=length(inputs(:,1));  % number of training patterns
epsi=10e-8;  % small number for preventing divide by zeros
eta=0.005;  % gradient step size for Pi
ncycle=60; % number of cycles of EM

alpha=0.1; % equivalent proportion of data represented by
% neibouring weights in within-modality topography

beta=16;  % inverse variance of between-modality
% interaction

gamma=72; % inverse variance of within-modality
% interaction

% calculate the within modality topology matrix

top=zeros(M);
for i=1:M
  for j=i:M
    top(i,j)=exp(-gamma*sqr(coord(i,K)-coord(j,K))/(2*T));;
  end;
end;
top=top+top';
rs=sum(top');
TOP=inv(eye(M)+2*alpha*(diag(rs)-top));
```

```
pi1=ones(M,1)/M;
pi2=ones(M,1)/M;
Pi=ones(M,M);
W1=rand(M,D);
W2=rand(M,D);
Wt1=rand(M,D);
Wt2=rand(M,D);

E=zeros(M*M,N);

P=zeros(M*M,N);
P1=zeros(M,N);
P2=zeros(M,N);
V1=0.5*ones(1,D);
V2=0.5*ones(1,D);
Pi=Pi/sum(sum(Pi));
I=[];
Cost=[];


for cycle=1:ncycle;
  T=1;

  % E step
  E=zeros(M*M,N);
  P=zeros(M*M,N);
  P1=zeros(M,N);
  P2=zeros(M,N);
  Costc=0;
  for l = 1:N
    x1=inputs(l,1:D);
    x2=inputs(l,D+1:2*D);
    for i=1:M
      for j=1:M
e=  -sqr((x1-W1(i,:))./sqrt(V1))/(2*T) ...
    -sqr((x2-W2(j,:))./sqrt(V2))/(2*T) ...
    -beta*sqr(coord(i,K)-coord(j,K))/(2*T) ...
    + log(pi1(i)) + log(pi2(j)) - log(Pi(i,j));
E((i-1)*M+j,l) = e;
P1(i,l)=P1(i,l)+exp(e);
```

```
P2(j,l)=P2(j,l)+exp(e);
      end;
    end;
    P(:,l)=exp(E(:,l));
    P(:,l)=P(:,l)/sum(P(:,l));
    for i=1:M*M
      Costc=Costc+E(i,l)*P(i,l);
    end;
    P1(:,l)=P1(:,l)/sum(P1(:,l));
    P2(:,l)=P2(:,l)/sum(P2(:,l));
  end;
  Cost=[Cost Costc];

  % M step

  % means

  for i=1:M
    Wt1(i,:)=P1(i,:)*inputs(:,1:D)/sum(P1(i,:));
    Wt2(i,:)=P2(i,:)*inputs(:,D+1:2*D)/sum(P2(i,:));
  end;

  W1=TOP*Wt1;
  W2=TOP*Wt2;


  % priors -- not updated in this model

  % pi1=sum(P1');
  % pi2=sum(P2');
  % pi1=pi1/sum(pi1);
  % pi2=pi2/sum(pi2);

  % diagonal variances

  V1=zeros(1,D);
  V2=zeros(1,D);
  for i=1:M
    for l=1:N
      V1=V1+P1(i,l)*(inputs(l,1:D)-W1(i,:)).^2;
      V2=V2+P2(i,l)*(inputs(l,D+1:2*D)-W2(i,:)).^2;
    end;
```

```
  end;
  V1=V1/N;
  V2=V2/N;

  % joint prior (joint entropy)

  Pest=reshape(sum(P'),M,M)'/N;
  for k=1:2;
    for i=1:M
      Pi(i,:)=pi1(i)*Pi(i,:)/sum(Pi(i,:));
    end;
    for i=1:M
      Pi(:,j)=pi2(j)*Pi(:,j)/sum(Pi(:,j));
    end;
  end;
  Pi=Pi-eta*(1./Pi).*Pest;
  pcut=0.01/M;
  Pi=(Pi>pcut).*Pi + (Pi<=pcut)*pcut;
  Pi=Pi/sum(sum(Pi));

  % calculate the mutual information from the joint entropy and normalize

  mi=mutinfo(Pest)/log(M);
  fprintf('cycle %g T %g V1 [%2.2f %2.2f] V2 [%2.2f %2.2f] lnL %g mi %g \n',...
      cycle,T,V1,V2,Costc,mi);
  I=[I mi];

  clear E P P1 P2;
  s=sprintf('save world%g.%g',sim,cycle); % save state at each step
  eval(s)
end;
```

Chapter 6

Conclusion

The first goal of this dissertation was to develop a computational framework for the study of sensorimotor integration and adaptation. The framework developed was based on the idea that information from multiple sources is integrated so as to obtain more accurate and reliable estimates of the state of the sensed world. This idea can be formalized within estimation theory, a branch of statistics, and leads to explicit models of sensorimotor integration. A testable premise of the computational framework is that adaptation to intersensory discrepancies is intimately linked to integration of information from multiple senses. From any model of multisensory integration, a model of intersensory adaptation can be derived that is consistent with the integration model.

What can be gained from developing such computational models in the first place? The models in this thesis have attempted to formalize intuitive ideas on the processes of integration and adaptation so as to make quantitative predictions possible. Like any scientific theory, a simple parsimonious model can both account for a large set of seemingly unrelated observations and provide an intuitive explanation for the phenomena being studied.

The computational models of integration and adaptation that were developed were tested through psychophysical experiments in three sensorimotor systems. The principal findings were

1. The patterns of visuo-auditory integration and adaptation suggest a principle of minimizing localization variance (Chapter 2).

2. The errors in estimating the location of the hand during a movement are also consistent with the minimum variance principle. Furthermore, the pattern of errors as a function of movement duration and external forces provides evidence for the existence of an internal model of the arm's dynamics in the CNS (Chapter 3).

3. The patterns of adaptation to local and contextual remappings of the visuomotor coordinate transformation suggest that it is represented with units which have large but localized receptive fields (Chapter 4).

Finally, the problem of converting information from several modalities into a common coordinate frame was examined (Chapter 5). Using a computational framework based on information theory, it was shown that mutually-aligned topographic maps can develop, in an unsupervised manner, from correlations between the inputs to different sensory modalities.

Bibliography

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

Abplanalp, P. and Held, R. (1965). Effects of de-correlated visual feedback on adaptation to wedge prisms. Paper presented at meetings of the Eastern Psychological Association, Atlantic City.

Ackley, D., Hinton, G., and Sejnowski, T. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, 9:147–169.

Albus, J. (1975). A new approach to manipulator control: The cerebellar model articulation controller (CMAC). *J. of Dynamic Systems, Measurement, and Control*, 97:220–227.

Andersen, R. (1987). Inferior parietal lobule function in spatial perception and visuomotor integration. In Plum, F. and Mountcastle, V., editors, *Handbook of Physiology: The Nervous System. Higher Functions of the Brain*, pages 483–518. Am. Physiol. Soc., Rockville, MD.

Aronson, E. and Rosenbloom, S. (1971). Space perception in early infancy: Perception within a common auditory-visual space. *Science*, 172:1161–1163.

Atkeson, C. (1989). Learning arm kinematics and dynamics. *Ann.Rev.Neurosci.*, 12:157–183.

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

Attneave, F. (1959). *Applications of Information Theory to Psychology.* University of Illinois Press, New York.

Baldi, P. and Hornik, K. (1989). Neural networks and principal components analysis: Learning from examples without local minima. *Neural Networks*, 2:53–58.

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

Barlow, H. (1989). Unsupervised learning. *Neural Computation*, 1:295–311.

Barlow, H. and Foldiak, P. (1989). Adaptation and decorrelation in the cortex. In *The Computing Neuron*, pages 54–72. Addison-Wesley.

Becker, S. (1992). *An information-theoretic unsupervised learning algorithm for neural networks*. Ph.D. Thesis. Dept. of Computer Science, U. Toronto, Toronto, Ontario.

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

Becker, W. and Fuchs, A. (1969). Further properties of the human saccadic system: eye movements and correction saccades with and without fixation points. *Vision Research*, 9:1247–1258.

Bedford, F. (1989). Constraints on learning new mappings between perceptual dimensions. *J. of Experimental Psychology: Human Perception and Performance*, 15(2):232–248.

Bedford, F. (1993). Perceptual and cognitive spatial learning. *J. of Experimental Psychology: Human Perception and Performance*, 19(3):517–530.

Blauert, J. (1983). *Spatial hearing: The Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA.

Braida, L. (1991). Crossmodal integration in the identification of consonant segments. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 43(3):647–678.

Brainard, M. and 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.

Bregman, A. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge, MA.

Broomhead, D. and Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2(3):321–355.

Bülthoff, H. and Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences (USA)*, 89:60–64.

Canon, L. (1970). Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *Journal of Experimental Psychology*, 84:141–147.

Carpenter, G. and Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *Computer*, pages 77–88.

Chambers, J. and Hastie, T. (1992). *Statistical Models in S.* Wadsworth, Pacific Grove, CA.

Choe, C., Welch, R., Guilford, R., and Juola, J. (1975). The "ventriloquist effect": Visual dominance or response bias? *Perception and Psychophysics*, 18:55–60.

Clark, B. and Graybiel, A. (1949). The effect of angular acceleration on sound localization: The audiogyral illusion. *Journal of Psychology*, 28:235–244.

Cohen, M. and Held, R. (1960). Degrading visual-motor coordination by exposure to disordered re-afferent stimulation. Paper presented at meetings of the Eastern Psychological Association, New York City.

Cottrell, G., Munro, P., and Zipser, D. (1987). Learning internal representations from gray-scale images: An example of extensional programming. In *Ninth Annual Conference of the Cognitive Science Society (Seattle 1987)*, pages 462–473. Erlbaum, Hillsdale.

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

Craig, J. (1986). *Introduction to Robotics.* Addison-Wesley, Reading, MA.

Craske, B. (1967). Adaptation to prisms: Change in internally registered eye-position. *British Journal of Psychology*, 58:329–335.

Cunningham, H. (1989). Aiming error under transformed spatial mappings suggests a structure for visual-motor maps. *J. of Experimental Psychology: Human Perception and Performance*, 15(3):493–506.

Day, R. and Wade, N. (1966). Visual spatial aftereffect from prolonged head-tilt. *Science*, 154:1201–1202.

de Sa, V. (1994). *Unsupervised Classification Learning from Cross-Modal Environmental Structure.* Ph.D. Thesis. Dept. of Computer Science, U. Rochester, Rochester, NY.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statistical Society Series B*, 39:1–38.

Dewar, R. (1970). Adaptation to displaced vision: Amount of optical displacement and practice. *Perception and Psychophysics*, 8:313–316.

Donaghue, J., Suner, S., and Sanes, J. (1990). Dynamic organization of primary motor cortex output to target muscles in adult rats. II. rapid reoganization following motor or mixed peripheral nerve lesions. *Experimental Brain Research*, 79:492–503.

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

Durbin, R. and Willshaw, D. (1987). An analogue approach to the travelling salesman problem using an elastic net method. *Nature*, 326(16):689–691.

Efstathiou, A. (1963). Correlated and de-correlated visual feedback in modifying eye-hand coordination. Paper presented at meetings of the Eastern Psychological Association, New York City.

Ewert, P. (1930). A study of the effect of inverted retinal stimulation upon spatially coordinated behaviour. *Genetic Psychology Monographs*, 7:242–244. Cited in Welch (1978) and Blauert (1983).

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

Fedderson, W., Sandel, T., Teas, D., and Jeffress, L. (1957). Localization of high frequency tones. *Journal of the Acoustical Society of America*, 29:988–991.

Fisher, G. (1960). Intersensory localisation in three modalities. *Bulletin British Psychological Society*, 41:24–25A.

Fisher, G. (1968). Agreement between the spatial senses. *Perceptual and Motor Skills*, 126:849–850.

Foldiak, P. (1990). Forming sparse representations by local anti-Hebbian learning. *Biological Cybernetics*, 64:165–170.

Freedman, S. and Pfaff, D. (1962). The effect of dichotic noise on auditory localization. *Journal of Auditory Research*, 2:305–310.

Freedman, S. and Zacks, J. (1964). Effects of active and passive movement upon auditory function during prolonged aptypical stimulation. *Perceptual and Motor Skills*, 18:361–366.

Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*, 20:121–136.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36:193–202.

Gallistel, C. (1980). *The Organization of Action: A New Synthesis.* Erlbaum, Hilladale, NJ.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Georgopoulos, A. (1990). Neurophysiology of reaching. In Jeannerod, M., editor, *Attention and performance XIII*, pages 227–263. Lawrence Erlbaum, Hillsdale.

Ghahramani, Z. (1995). Factorial learning and the *EM* algorithm. In Tesauro, G., Touretzky, D., and Leen, T., editors, *Advances in Neural Information Processing Systems 7*. MIT Press, Cambridge, MA.

Ghez, C., Gordon, J., and Ghilardi, M. (1995). Impairments of reaching movements in patients without proprioception II: Effects of visual information on accuracy. *J. Neurophysiol.*, 73(1):361–372.

Ghez, C., Gordon, J., Ghilardi, M., Christakos, C., and Cooper, S. (1990). Roles of proprioceptive input in the programming of arm trajectories. *Cold Spring Harbor Symp.Quant.Biol*, 55:837–847.

Goldstein, K. and Rosenthal-Veit, O. (1926). Uber akustische lokalisation une deren beeinflussbarkeit durch andere sinnesreize. *Psychol. Forsch.*, 8:318–335.

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

Gordon, J., Ghilardi, M., and Ghez, C. (1995). Impairments of reaching movements in patients without proprioception I: Spatial errors. *J. Neurophysiol.*, 73(1):347–360.

Graybiel, A. and Niven, J. (1951). the effect of a change in direction of resultant force on sound localization: The audiogravic illusion. *J. Exper. Psychol.*, 42:227–230.

Grossberg, S. (1987). Competitive Learning: From interactive activation to adaptive resonance. *Cognitive Science*, 11:23–63.

Guitton, D. and 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 . *Journal of Neurophysiology*, 66:1605–1623.

Harris, C. (1965). Perceptual adaptation to inverted, reversed, and displaced vision. *Psychological Review*, 72:419–444.

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

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

Hay, J. and Pick, H. (1966). Gaze-contingent prism adaptation: Optical and motor factors. *J. of Experimental Psychology*, 72:640–648.

Hay, J., Pick, H., and Ikeda, K. (1965). Visual capture produced by prism spectacles. *Psychonomic Science*, 2:215–216.

Hebb, D. (1949). *The Organization of Behavior.* Wiley, New York.

Held, R. (1955). Shifts in binaural localization after prolonged exposure to atypical combinations of stimuli. *Amer. J. of Psych.*, 68:526–548.

Held, R. (1962). Adaptation to rearrangement and visual-spatial aftereffects. *Psychologische Beitrage*, 6:439–450.

Held, R., Efstathiou, A., and Greene, M. (1966). Adaptation to displaced and delayed visual feedback from the hand. *J. Exp. Psychol.*, 72:887–891.

Held, R. and Hein, A. (1958). Adaptation to disarranged hand-eye coordination contingent upon re-afferent stimulation. *Perceptual and Motor Skills*, 8:97–90.

Henson, D. (1978). Corrective saccades: effects of altering visual feedback. *Vision Research*, 18:63–67.

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

Hinton, G. and Zemel, R. (1994). Autoencoders, minimum description length, and Helmholtz free energy. In Cowan, J., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems 6.* Morgan Kaufmanm Publishers, San Francisco, CA.

Howard, I. (1982). *Human visual orientation.* Wiley, Chichester, England.

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

Imamizu, H., Uno, Y., and Kawato, M. (1994). Internal representations of the motor apparatus: Implication from generalization in visuo-motor learning. *Unpublished Manuscript.*

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

Jack, C. and Thurlow, W. (1973). Effects of degree of association and angle of displacement on the "ventriloquism" effect. *Perceptual and Motor Skills*, 37:967–979.

Jacobs, R., Jordan, M., Nowlan, S., and Hinton, G. (1991). Adaptive mixture of local experts. *Neural Computation*, 3:79–87.

Jay, M. and Sparks, D. (1984). Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature*, 309:345–347.

Johnson, R. and Wichern, D. (1992). *Applied Multivariate Statistical Analysis*. Prentice Hall, Englewood Cliffs, NJ.

Jones, B. and Kabanoff, B. (1975). Eye movements in auditory space perception. *Percept. and Psychophys.*, 17:241–245.

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

Jordan, M. I. (1995). Computational aspects of motor control and motor learning. In Heuer, H. and Keele, S., editors, *Handbook of Perception and Action: Motor Skills*. Academic Press, New York.

Juang, B. and Rabiner, L. (1991). Hidden Markov models for speech recognition. *Technometrics*, 33:251–272.

Julesz, B. (1971). *Foundations of Cyclopean Perception*. University of Chicago Press, Chicago, IL.

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

Kandel, E., Schwartz, J., and T.M. Jessel, e. (1991). *Principles of Neural Science*. Elsevier, New York.

Kaufman, L. (1974). *Sight and Mind*. Oxford University Press, New York.

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

Klemm, O. (1909). Localization of sensory impressions in connection with disparate additional stimuli (in german). *Psychol. Stud.*, 5:73–162. Cited in Blauert (1983).

Klemm, O. (1918). Investigations of the localization of sound stimuli III: On what is contributed by two-eared hearing (in german). *Arch. ges. Psychol.*, 38:71–114. Cited in Blauert (1983).

Knudsen, E. (1982). Auditory and visual maps of space in the optic tectum of the owl. *J. Neurosci.*, 2:1177–1194.

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

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

Koh, K. and Meyer, D. (1991). Function learning: Induction of continuous stimulus-response relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(5):811–836.

Kohler, I. (1950). Development and alterations of the perceptual world: conditioned sensations. *Proceedings of the Austrian Academy of Sciences*, 227.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69.

Kohonen, T. (1989). *Self-organization and associative memory.* Springer-Verlag, Berlin.

Konishi, M., Takahashi, T., Wagner, H., Sullivan, W., and Carr, C. (1988). Neurophysiological and anatomical substrates of sound localization in the owl. In Edelman, G., Gall, W., and Cowan, W., editors, *Auditory Function: Neurobiological Bases of Hearing*, pages 721–745. Wiley, New York.

Kravitz, J. (1972). Conditioned adaptation to prismatic displacement. *Perception and Psychophysics*, 11:38–42.

Kravitz, J. and Yaffe, F. (1972). Conditioned adaptation to prismatic displacement with a tone as the conditional stimulus. *Perception and Psychophysics*, 12:305–308.

Lackner, J. (1973). Visual rearrangement affects auditory localization. *Neuropsychologia*, 11:29–32.

Lackner, J. (1974a). Changes in auditory localization during body tilt. *Acta Otolaryngologica (Stockholm)*, pages 19–28.

Lackner, J. (1974b). Influence of visual rearrangement and visual motion on sound localization. *Neuropsychologia*, 12:291–293.

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

Linsker, R. (1986b). From basic network principles to neural architecture: Emergence of orientation columns. *Proceedings of the National Academy of Sciences USA*, 83:8779–8783.

Linsker, R. (1986c). From basic network principles to neural architecture: Emergence of spatial opponent cells. *Proceedings of the National Academy of Sciences USA*, 83:7508–7512.

Linsker, R. (1988). Self-organization in a perceptual network. *IEEE Computer*, 21:105–117.

Luttrell, S. (1989). Self-organisation: a derivation from first principles of a class of learning algorithms. In *IJCNN: International Joint Conference on Neural Networks*, volume 2, pages 495–498. IEEE, New York.

Luttrell, S. (1994). A Bayesian analysis of self-organizing maps. *Neural Computation*, 6(5):767–794.

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

Matin, L. (1986). Visual localization and eye movements. In Boff, K., Kaufman, L., and Thomas, J., editors, *Handbook of Perception and Human Performance. Volume I: Sensory Processes and Perception*. Wiley.

Maunsell, J., Nealy, T., Sclar, G., and DePriest, D. (1989). Representation of extraretinal information in monkey visual cortex. In Lam, D. M.-K. and Gilbert, C., editors, *Neural mechanisms of visual perception. Proceedings of the Retina Rearch Foundation Symposia, Volume Two*, pages 223–236. Gulf Publishing, Houston.

McGonigle, B. and Flook, J. (1978). Long-term retention of single and multistate prismatic adaptation by humans. *Nature*, 272:364–366.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264:746–748.

McLachlan, G. and Basford, K. (1988). *Mixture models: Inference and applications to clustering*. Marcel Dekker, New York.

Merzenich, M., Kaas, J., Wall, J., Nelson, R., Sur, M., and Felleman, D. (1983). Topographic reorganization of somatosensory cortical areas 3b and 1 in adult monkeys following restricted deafferentation. *Neuroscience*, 8:33–55.

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

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

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

Moody, J. and Darken, C. (1989). Fast learning in networks of locally-tuned processing units. *Neural Computation*, 1(2):281–294.

Morrell, F. (1972). Visual system's view of acoustic space. *Nature*, 238:44–46.

Mountcastle, V., Lynch, J., Georgopoulos, A., Sakata, H., and Acuna, C. (1975). Posterior parietal association cortex of the monkey: command function for operations within extrapersonal space. *J. Neurophysiol.*, 38:871–908.

Neal, R. (1992). Connectionist learning of belief networks. *Artificial Intelligence*, 56:71–113.

Nowlan, S. (1991). *Soft Competitive Adaptation: Neural Network Learning Algorithms based on Fitting Statistical Mixtures*. Ph.D. Thesis CMU-CS-91-126. School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *J. of Mathematical Biology*, 15(3):267–273.

Oja, E. (1989). Neural networks, principal components, and subspaces. *International Journal of Neural Systems*, 1(1):61–68.

Parisi, G. (1988). *Statistical Field Theory*. Addison-Wesley, Redwood City, CA.

Pick, H., Warren, D., and Hay, J. (1969). Sensory conflict in judgements of spatial direction. *Percept. and Psychophys.*, 6(4).

Plumbley, M. (1991). *On Information Theory and Unsupervised Neural Networks*. Cambridge University Engineering Department CUED/F-INFENG/TR.78, Cambrige, UK.

Plumbley, M. and Fallside, F. (1988). An information-theoretic approach to unsupervised learning. In Touretzky, D., Hinton, G., and Sejnowski, T., editors, *Proceedings of the 1988 Connectionist Models Summer School*, pages 239–245. Morgan-Kaufmann, San Mateo, CA.

Poggio, T. (1990). A theory of how the brain might work. *Cold Spring Harbor Symposium on Quantitative Biology*, pages 899–910.

Poggio, T. and Girosi, F. (1989). A theory of networks for approximation and learning. *AI Lab Memo 1140, MIT*.

Poggio, T. and Hurlbert, A. (1994). Observations on cortical mechanisms for object recognition and learning. In Koch, C. and Davis, J., editors, *Large-Scale Neuronal Theories of the Brain*, pages 153–182. MIT Press, Cambridge, MA.

Pöppel, E. (1973). Comment on "visual system's view of acoustic space". *Nature*, 243:231.

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

Pouget, A. and Sejnowski, T. (1995). Spatial representation in the parietal cortex may use basis functions. In Tesauro, G., Touretzky, D., and Leen, T., editors, *Advances in Neural Information Processing Systems 7*. MIT Press, Cambridge, MA.

Prablanc, C., Tzavaras, A., and Jeannerod, M. (1975). Adaptation of the two arms to opposite prism displacements. *Quarterly Journal of Experimental Psychology*, 27:667–671.

Press, W., Flannery, B., Teukolsky, S., and Vetterling, W. (1988). *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK.

Radeau, M. and Bertelson, P. (1974). The after-effects of ventriloquism. *Quarterly Journal of Experimental Psychology*, 26:63–71.

Rayleigh, L. (1907). On our perception of sound direction. *Philos. Mag.*, 13:214–232. Cited in Middlebrooks & Green (1991).

Recanzone, G., Merzenich, M., Jenkins, W., Grajski, K., and Dinse, H. (1992). Topographic reorganization of the hand representation in cortical area 3b of owl monkeys trained on a frequency discrimination task. *J. Neurophysiology*, 67:1031–1056.

Redding, G. and Wallace, B. (1988). Adaptive mechanisms in perceptual-motor coordination: Components of prism adaptation. *Journal of Motor Behavior*, 20:242–254.

Redding, G. and Wallace, B. (1993). Adaptive coordination and alignment of eye and hand. *Journal of Motor Behavior*, 25:75–88.

Rissanen, J. (1989). *Stochastic Complexity in Statistical Inquiry*. World Scientific, Singapore.

Robinson, D. (1987). Control of eye movements. In Plum, F. and Mountcastle, V., editors, *Handbook of Physiology: The Nervous System II*, pages 1275–1320. Am. Physiol. Soc., Rockville, MD.

Robinson, D., Gordon, J., and Gordon, S. (1986). A model of the smooth pursuit eye movement system. *Biol.Cybern.*, 55:43–57.

Roe, A., Pallas, S., Hahm, J.-O., and Sur, M. (1990). A map of visual space Induced in primary auditory cortex. *Science*, 250.

Rosenbaum, D., Engelbrecht, S., Bushe, M., and Loukopoulos, L. (1993). Knowledge model for selecting and producing reaching movements. *Journal of Motor Behavior*, 25(3):217–227.

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

Rumelhart, D. and Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science*, 9:75–112.

Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O., Lu, S.-T., and Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, 127(1):141–145.

Sanes, J., Suner, S., and Donaghue, J. (1990). Dynamic organization of primary motor cortex output to target muscles in adult rats. I. long-term patterns of reoganization following motor or mixed peripheral nerve lesions. *Experimental Brain Research*, 79:479–491.

Sanger, T. (1989). Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2:459–473.

Sanger, T. (1994). Using the genereralization properties of a neural network to infer the number and shape of its hidden units. *Unpublished manuscript*.

Scharf, B. and Houtsma, A. (1986). Audition II: Loudness, pitch, localization, aural distortion, pathology. In Boff, K., Kaufman, L., and Thomas, J., editors, *Handbook of Perception and Human Performance. Volume I: Sensory Processes and Perception*. Wiley.

Sekiyama, K. and Tohkura, Y. (1991). McGurk effect in non-English listerners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *J. Acoust. Soc. Am.*, 90(4):1797–1805.

Sekuler, R. and Blake, R. (1990). *Perception (2nd Edition)*. McGraw-Hill, New York.

Shadmehr, R. and Mussa-Ivaldi, F. (1994). Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.*, 14(5):3208–3224.

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

Shannon, C. and Weaver, W. (1949). *The Mathematical Theory of Communication.* University of Illinois Press, Urbana, IL.

Shelhamer, M., Robinson, D., and Tan, H. (1991). Context-specific gain switching in the human vestibuloocular reflex. In Cohen, B., Tomko, D., and Guedry, F., editors, *Annals Of The New York Academy Of Sciences*, volume 656, pages 889–891. New York Academy Of Sciences, New York.

Shepard, R. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89:305–333.

Shinn-Cunningham, B. (1994). *Adaptation to Supernormal Auditory Localization Cues in an Auditory Virtual Environment.* Ph.D. Thesis, Dept. of Electrical Engineering and Computer Science, MIT, Cambridge, MA.

Snyder, L., Brotchie, P., and Andersen, R. (1993). World-centered encoding of location in posterior parietal cortex of monkey. *Society For Neuroscience Abstracts.*, 19((1-3)):770.

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

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

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

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

Stevens, S. and Newman, E. (1936). The localization of actual sources of sound. *American Journal of Psychology*, 48:297–306.

Stratton, G. (1897a). Upright vision and the retinal image. *Psychological Review*, 4:182–187. Cited in Welch (1978)and Blauert (1983).

Stratton, G. (1897b). Vision without inversion of the retinal image. *Psychological Review*, 4:341–360, 463–481. Cited in Blauert (1983) and Welch & Warren (1986).

Sutton, R. and Barto, A. (1981). Toward a modern theory of adaptive networks: expectation and prediction. *Psychol.Rev.*, 88:135–170.

Tikhonov, A. and Arsenin, V. (1977). *Solutions of Ill-Posed Problems.* W.H. Winston, Washington, D.C.

Titterington, D., Smith, A., and Makov, U. (1985). *Statistical analysis of finite mixture distributions*. Wiley, Chichester.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in striate cortex. *Kybernetik*, 14:85–100.

von Helmholtz, H. (1925). *Treatise on physiological optics (1867)*. Optical Society of America, Rochester, New York.

Warren, D. and Pick, H. (1970). Intermodality relations in blind and sighted people. *Perception and Psychophysics*, 8:430–432.

Welch, R. (1971). Discriminative conditioning of prism adaptation. *Perception and Psychophysics*, 10:90–92.

Welch, R. (1978). *Perceptual Modification*. Academic Press, New York.

Welch, R. (1986). Adaptation to space perception. In Boff, K., Kaufman, L., and Thomas, J., editors, *Handbook of perception and performance*, volume 1, pages 24–1–24–45. Wiley–Interscience, New York.

Welch, R., Bridgeman, B., Anand, S., and Browman, K. (1993). Alternating prism exposure causes dual adaptation and generalization to a novel displacement. *Perception and Psychophysics*, 54(2):195–204.

Welch, R. and Warren, D. (1986). Intersensory Interactions. In Boff, K., Kaufman, L., and Thomas, J., editors, *Handbook of Perception and Human Performance. Volume I: Sensory Processes and Perception*. Wiley.

Wertheimer, M. (1961). Psychomotor co-ordination of auditory-visual space at birth. *Science*, 134:1692.

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

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

Widrow, B. and Hoff, M. (1960). Adaptive switching circuits. In *IRE WESCON Convention Record*, volume 4, pages 96–104. IRE, New York.

Wiesel, T. and Hubel, D. (1963). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *J. Neurophysiology*, 26:994–1002.

Wilson, H., Levi, D., Maffei, L., Rovamo, J., and DeValois, R. (1990). The perception of form: Retina to striate cortex. In Spillman, L. and Werner, J., editors, *Visual Perception: The Neurophysiological Foundations*, pages 231–272. Academic Press, San Diego.

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

Wooster, M. (1923). Certain factors in the development of a new spatial coordination. *Psychological Monographs*, 32(4). Cited in Welch (1978).

Yuille, A. (1990). Generalized deformable models, statistical physics, and matching problems. *Neural Computation*, 2(1):1–24.

Zemel, R. (1993). *A minimum description length framework for unsupervised learning*. Ph.D. Thesis, Dept. of Computer Science, University of Toronto, Toronto, Canada.

Zipser, D. and Andersen, R. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331:679–684.