

UMIACS-TR-97-56  
CS-TR-3816

August 1997

# **A Motor Control Model Based on Self-organizing Feature Maps**

Yinong Chen

Institute for Advanced Computer Studies  
Department of Computer Science  
University of Maryland  
College Park, MD 20742

**Acknowledgement:** This work was supported by NIH awards NS29414.



## Abstract

Title of Dissertation: A Motor Control Model Based on Self-organizing  
Feature Maps

Yinong Chen, Doctor of Philosophy, 1997

Dissertation directed by: Professor James A. Reggia  
Department of Computer Science

Self-organizing feature maps have become important neural modeling methods over the last several years. These methods have not only shown great potential in application fields such as motor control, pattern recognition, optimization, etc, but have also provided insights into how mammalian brains are organized. Most past work developing self-organizing features maps has focused on systems with a single map that is solely sensory in nature. This research develops and studies a model which has multiple self-organizing feature maps in a closed-loop control system, and that involves motor output as well as proprioceptive and/or visual sensory input. The model is driven by a simulated arm that moves in 3D space.

By applying initial activations at randomly selected motor cortex regions, the neural network model spontaneously self-organizes, and demonstrates the appearance of multiple, reasonably stable motor and proprioceptive sensory maps and their interrelationships to each other. These cortical feature maps capture the mechanical constraints imposed by the model arm. They are aligned in a way consistent with a *temporal correlation hypothesis*: temporally correlated features usually cause their corresponding cortical map representations to be spatially correlated.

Simulations of variations of the motor control model with visual inputs indicates the formation of visual input maps. These maps are also partially aligned with motor output maps, reflecting the degree of temporal correlations during training. The simultaneous presence of proprioceptive input causes the visual input maps to distinguish pairs of antagonist muscles and to be correlated with only one muscle in each pair. Moreover, some theoretical analysis with a simplified model gives insights into the nature of cortical feature maps and sheds light on the driving force behind map correlations. All of these results have provide more understanding about the organization of cortical feature maps, and how these maps might be used to achieve consistent motor commands based on sensory feedback.

# **A Motor Control Model Based on Self-organizing Feature Maps**

by

Yinong Chen

Dissertation submitted to the Faculty of the Graduate School  
of the University of Maryland in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
1997

Advisory Committee:

Professor James A. Reggia, Chairman/Advisor  
Associate Professor Don Perlis  
Associate Professor Christos Faloutsos  
Associate Professor Yun Peng  
Associate Professor Avis Cohen, Dean's representative

# Contents

<b>List of Tables</b>	<b>iv</b>
<b>List of Figures</b>	<b>vi</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Background</b>	<b>4</b>
2.1 General Information on Self-Organizing Maps . . . . .	4
2.1.1 Using Maps for Computational Purposes . . . . .	5
2.1.2 Biological Modeling Using Maps . . . . .	6
2.2 Cortical Lateral Inhibition . . . . .	8
2.3 Feature Maps for Motor Control . . . . .	10
2.3.1 Modeling of Motor Cortex: Visuo-motor-coordination . . . . .	10
2.3.2 Proprioceptive Cortex Map . . . . .	13
<b>3 The Motor Control Model</b>	<b>16</b>
3.1 The Arm Model . . . . .	16
3.2 Network Structure . . . . .	19
3.3 Activation and Learning Rules . . . . .	21
3.4 Experimental Methods . . . . .	22
<b>4 Motor Control Model with Proprioceptive Input Only</b>	<b>25</b>
4.1 The Model and Parameters . . . . .	25
4.2 Results . . . . .	26
4.2.1 Input Map Formation and Characteristics . . . . .	27
4.2.2 Motor Output Map Formation and Characteristics . . . . .	35
4.2.3 Consistency Between Input and Output Maps . . . . .	35
4.2.4 Elements Tuned to Multiple Features . . . . .	39
4.2.5 Sensitivity to Simultaneous Weight Adaptation . . . . .	40
4.2.6 Lesioning Study of the Motor Control Model . . . . .	41
4.3 Discussion . . . . .	41
<b>5 Motor Control Model with Only Visual Input</b>	<b>44</b>
5.1 The Model . . . . .	44
5.2 Experimental Methods . . . . .	46
5.3 Results . . . . .	48

5.3.1	Cortical Map Formation in MI . . . . .	48
5.3.2	Correlations Between the MI Input and Output Maps . . . . .	50
5.4	Discussion . . . . .	56
<b>6</b>	<b>Motor Control Model with Combined Proprioceptive and Visual Inputs</b>	<b>58</b>
6.1	Experimental Methods . . . . .	58
6.2	Results . . . . .	60
6.2.1	PI Input Maps with Respect to the Proprioceptive Inputs . . . . .	60
6.2.2	MI Proprioceptive Input Maps . . . . .	63
6.2.3	MI Output Maps and Their Relation to Proprioceptive Input Maps . . . . .	65
6.2.4	MI Visual Input Maps and Their Relation to MI Output Maps . . . . .	68
6.2.5	The Relationship between Proprioceptive Input Maps and Visual Input Maps in MI . . . . .	74
6.3	Discussion . . . . .	75
<b>7</b>	<b>Some Analysis of Cortical Feature Map Formation</b>	<b>77</b>
7.1	The Abstract Model . . . . .	77
7.2	Activation Patterns of Output Units . . . . .	78
7.3	The Relationship of Cortical Maps with Perfectly Correlated Features . . . . .	87
7.4	The Relationship of Cortical Maps with Perfectly Anti-correlated Features . . . . .	91
<b>8</b>	<b>Conclusion</b>	<b>97</b>
8.1	Summary and Contributions . . . . .	97
8.2	Limitations and Future Work . . . . .	99
<b>A</b>	<b>Complete Cortical Feature Maps</b>	<b>101</b>
A.1	Cortical Feature Maps of Motor Control Model with Proprioceptive Input Only . . . . .	101
A.1.1	Proprioceptive input maps in PI layer . . . . .	101
A.1.2	Proprioceptive input maps in MI layer . . . . .	108
A.1.3	Motor output maps in MI layer . . . . .	114
A.2	Cortical Feature Maps of Motor Control Model with Visual Input Only . . . . .	117
A.2.1	Motor output maps in MI layer . . . . .	117
A.2.2	Visual input maps in MI layer . . . . .	120
A.3	Cortical Feature Maps of Motor Control Model with Combined Proprioceptive Input and Visual Input . . . . .	123
A.3.1	Proprioceptive input maps in PI layer . . . . .	123
A.3.2	Proprioceptive input maps in MI layer . . . . .	129
A.3.3	Motor output maps in MI layer . . . . .	135
A.3.4	Visual input maps in MI layer . . . . .	138
<b>B</b>	<b>Lesioning the Motor Control Model</b>	<b>141</b>
B.1	Unlesioned Model . . . . .	141
B.2	Focal Lesions in Proprioceptive Cortex . . . . .	143
B.3	Comments . . . . .	146

## List of Tables

4.1	Parameters used in activation update rule. . . . .	26
4.2	Parameters used in activation dispersal rule and learning rule. . . . .	26
4.3	Labeling of muscle length and tension in illustrations. . . . .	26
4.4	Similarity values between length and tension input features in proprioceptive cortex before training. . . . .	31
4.5	Similarity values between length and tension input features in proprioceptive cortex after training. Those values that are bigger than 0.7 are indicated in bold. . . . .	31
4.6	Similarity values between length and tension proprioceptive input features in the motor cortex layer before (left) and after (right) training. Those values that are bigger than 0.7 are indicated in bold. . . . .	35
4.7	Similarity values between length and tension input features in motor cortex before (left) and after (right) training. Those values that are bigger than 0.7 are in bold style. . . . .	38
4.8	Numbers of implausibly tuned PI and MI layer elements (threshold=0.5). . . . .	39
5.1	Parameters used in activation update rule. . . . .	48
5.2	Parameters used in activation dispersal rule and learning rule. . . . .	48
5.3	Similarity values between visual input and motor output features before training. . . . .	52
5.4	Similarity values between visual input and motor output features after training. Those values that are bigger than the biggest values in Table 5.3 are in bold style. . . . .	54
5.5	Similarity values between visual input and motor output features after training in four other simulations with different initial weights and training patterns. Those values that are bigger than the biggest values in Table 5.3 are in bold style. . . . .	56
6.1	Parameters used in activation update rule. . . . .	60
6.2	Parameters used in activation dispersal rule and learning rule. . . . .	60
6.3	Similarity values between length and tension input features in proprioceptive cortex before (left) and after (right) training. Those values that are bigger than 0.7 are in bold style. . . . .	63
6.4	Similarity values between length and tension input features in motor cortex before (left) and after (right) training. Those values that are bigger than 0.7 are in bold style. . . . .	65
6.5	Similarity values between length and tension input features in motor cortex before (left) and after (right) training. Those values that are bigger than 0.7 are in bold style. . . . .	68

6.6	Similarity values between visual input and motor output features before (left) and after (right) training. Those values that are bigger than 0.6 are in bold style. . . . .	71
6.7	Similarity values between visual input and motor output features after training in four other simulations with different initial weights and training patterns. Those values that are bigger than 0.7 are in bold style. . . . .	73
6.8	Similarity values between visual input and length (left) and tension (right) proprioceptive input features after training. Those values that are bigger than 0.5 are in bold style. . . . .	74



## List of Figures

2.1	Topographic maps in human somatosensory and motor cortex . . . . .	5
2.2	“Mexican Hat” like activation pattern . . . . .	9
2.3	A schematic diagram of lateral connections to produce central excitatory, peristimulus inhibitory activation patterns. Other implementation may have the same effect.	10
2.4	Competitive activation mechanism . . . . .	11
2.5	Network architecture . . . . .	14
3.1	Schematic diagram of the closed loop motor system . . . . .	17
3.2	Model arm . . . . .	18
3.3	Generic joint . . . . .	19
3.4	Network architecture of the motor control model . . . . .	20
4.1	Sensitivity of PI elements to muscle length . . . . .	27
4.2	Tuning of PI elements to the length of the upper arm extensor (E) and flexor (F) before (left) and after (right) training (threshold=0.4). . . . .	28
4.3	PI elements tuned to length and tension of upper arm extensor and flexor . . . . .	30
4.4	MI elements tuned to length and tension of upper arm extensor and flexor . . . . .	32
4.5	Intersection of activation areas. . . . .	33
4.6	Representative distributions of incoming weight vector in MI . . . . .	34
4.7	MI output map before (left) and after (right) training for upper arm extensor (E) and flexor (F) (threshold=0.4). . . . .	36
4.8	Comparing post-training MI output maps with MI input maps . . . . .	37
4.9	Map of MI elements strongly activating multiple muscles (threshold=0.4) . . . . .	40
5.1	The motor system with visual inputs . . . . .	45
5.2	The coordinate system in the arm movement space relative to a human body . . . . .	46
5.3	The tuning curves of the three visual elements coding hand positions in the X-dimension. . . . .	47
5.4	The MI output maps before (left) and after (right) training. Only maps for the upper arm extensor (E) and flexor (F) are shown (threshold=0.4). . . . .	49
5.5	The MI input maps with respect to visual input (in the X dimension), before (left) and after (right) training. X1, X2 and X3 code the negative, middle and positive range in the X dimension. . . . .	51
5.6	a. MI output map for upper arm extensor (E) after training (threshold=0.7); b. MI input map for negative range of the Y dimension (Y3) from visual input (threshold=0.3). . . . .	53

5.7	A Density Plot of the similarity table . . . . .	55
6.1	The motor control system with both proprioceptive and visual inputs . . . . .	59
6.2	The PI input maps before (left) and after (right) training. Only length maps for the upper arm extensor (E) and flexor (F) are shown (threshold=0.2). . . . .	61
6.3	Comparison of PI input maps with respect to length (left) and tension (right) after training. Only maps for the upper arm extensor (E,e) and flexor (F,f) are shown (threshold=0.2). . . . .	62
6.4	Comparison of MI input maps with respect to length (left) and tension (right) after training. Only maps for the upper arm extensor (E,e) and flexor (F,f) are shown (threshold=0.4). . . . .	64
6.5	The MI output maps before (left) and after (right) training. Only maps for the upper arm extensor (E,e) are shown (threshold=0.4). . . . .	66
6.6	The comparison of the MI input maps with respect to the length (a) and tension (b) of upper arm extensor, and the MI output maps of upper arm extensor (c) and flexor (d) after training (threshold=0.4). . . . .	67
6.7	The MI visual input maps before (left) and after (right) training. Only the MI input maps for the X dimension are shown (threshold=0.3). . . . .	69
6.8	Comparison of MI motor output map with MI input map after training . . . . .	70
6.9	A Density Plot of the similarity Table . . . . .	72
7.1	The abstract model . . . . .	78
7.2	Drawing of $L_1 = 0$ and $L_2 = 0$ in the two dimensional space $(a_1, a_2)$ . . . . .	82
7.3	Drawing of $L_i = 0$ and $L_k = 0$ in subspace $(a_i, a_k)$ of the n-dimensional space . . . . .	84
7.4	A subspace (in dimension $i$ and $j$ ) of the $m$ dimensional input space . . . . .	91
7.5	An example of a small network . . . . .	96
A.1	Tuning of PI elements to the length of the upper arm extensor (E) and flexor (F) before (left) and after (right) training (threshold=0.4). . . . .	102
A.2	Tuning of PI elements to the length of the upper arm abductor (B) and adductor (D) before (left) and after (right) training (threshold=0.4). . . . .	103
A.3	Tuning of PI elements to the length of the lower arm extensor or opener (O) and flexor or closer (C) before (left) and after (right) training (threshold=0.4). . . . .	104
A.4	Tuning of PI elements to the tension of the upper arm extensor (e) and flexor (f) before (left) and after (right) training (threshold=0.4). . . . .	105
A.5	Tuning of PI elements to the tension of the upper arm abductor (b) and adductor (d) before (left) and after (right) training (threshold=0.4). . . . .	106
A.6	Tuning of PI elements to the tension of the lower arm extensor or opener (o) and flexor or closer (c) before (left) and after (right) training (threshold=0.4). . . . .	107
A.7	Tuning of MI elements to the length of the upper arm extensor (E) and flexor (F) before (left) and after (right) training (threshold=0.4). . . . .	108
A.8	Tuning of MI elements to the length of the upper arm abductor (B) and adductor (D) before (left) and after (right) training (threshold=0.4). . . . .	109
A.9	Tuning of MI elements to the length of the lower arm extensor or opener (O) and flexor or closer (C) before (left) and after (right) training (threshold=0.4). . . . .	110

A.10	Tuning of MI elements to the tension of the upper arm extensor (e) and flexor (f) before (left) and after (right) training (threshold=0.4). . . . .	111
A.11	Tuning of MI elements to the tension of the upper arm abductor (b) and adductor (d) before (left) and after (right) training (threshold=0.4). . . . .	112
A.12	Tuning of MI elements to the tension of the lower arm extensor or opener (o) and flexor or closer (c) before (left) and after (right) training (threshold=0.4). . . . .	113
A.13	MI output map before (left) and after (right) training for upper arm extensor (E) and flexor (F) (threshold=0.4). . . . .	114
A.14	MI output map before (left) and after (right) training for upper arm abductor (B) and adductor (D) (threshold=0.4). . . . .	115
A.15	MI output map before (left) and after (right) training for lower arm extensor or opener (O) and flexor or closer (C) (threshold=0.4). . . . .	116
A.16	MI output map before (left) and after (right) training for upper arm extensor (E) and flexor (F) (threshold=0.4). . . . .	117
A.17	MI output map before (left) and after (right) training for upper arm abductor (B) and adductor (D) (threshold=0.4). . . . .	118
A.18	MI output map before (left) and after (right) training for lower arm extensor or opener (O) and flexor or closer (C) (threshold=0.4). . . . .	119
A.19	The MI input maps with respect to visual input (in the X dimension), before (left) and after (right) training. X1, X2 and X3 code the negative, middle and positive range in the X dimension (threshold=0.3). . . . .	120
A.20	The MI input maps with respect to visual input (in the Y dimension), before (left) and after (right) training. Y1, Y2 and Y3 code the negative, middle and positive range in the Y dimension (threshold=0.3). . . . .	121
A.21	The MI input maps with respect to visual input (in the Z dimension), before (left) and after (right) training. Z1, Z2 and Z3 code the negative, middle and positive range in the Z dimension (threshold=0.3). . . . .	122
A.22	Tuning of PI elements to the length of the upper arm extensor (E) and flexor (F) before (left) and after (right) training (threshold=0.2). . . . .	123
A.23	Tuning of PI elements to the length of the upper arm abductor (B) and adductor (D) before (left) and after (right) training (threshold=0.2). . . . .	124
A.24	Tuning of PI elements to the length of the lower arm extensor or opener (O) and flexor or closer (C) before (left) and after (right) training (threshold=0.2). . . . .	125
A.25	Tuning of PI elements to the tension of the upper arm extensor (e) and flexor (f) before (left) and after (right) training (threshold=0.2). . . . .	126
A.26	Tuning of PI elements to the tension of the upper arm abductor (b) and adductor (d) before (left) and after (right) training (threshold=0.2). . . . .	127
A.27	Tuning of PI elements to the tension of the lower arm extensor or opener (o) and flexor or closer (c) before (left) and after (right) training (threshold=0.2). . . . .	128
A.28	Tuning of MI elements to the length of the upper arm extensor (E) and flexor (F) before (left) and after (right) training (threshold=0.4). . . . .	129
A.29	Tuning of MI elements to the length of the upper arm abductor (B) and adductor (D) before (left) and after (right) training (threshold=0.4). . . . .	130
A.30	Tuning of MI elements to the length of the lower arm extensor or opener (O) and flexor or closer (C) before (left) and after (right) training (threshold=0.4). . . . .	131

A.31	Tuning of MI elements to the tension of the upper arm extensor (e) and flexor (f) before (left) and after (right) training (threshold=0.4). . . . .	132
A.32	Tuning of MI elements to the tension of the upper arm abductor (b) and adductor (d) before (left) and after (right) training (threshold=0.4). . . . .	133
A.33	Tuning of MI elements to the tension of the lower arm extensor or opener (o) and flexor or closer (c) before (left) and after (right) training (threshold=0.4). . . . .	134
A.34	MI output map before (left) and after (right) training for upper arm extensor (E) and flexor (F) (threshold=0.4). . . . .	135
A.35	MI output map before (left) and after (right) training for upper arm abductor (B) and adductor (D) (threshold=0.4). . . . .	136
A.36	MI output map before (left) and after (right) training for lower arm extensor or opener (O) and flexor or closer (C) (threshold=0.4). . . . .	137
A.37	The MI input maps with respect to visual input (in the X dimension), before (left) and after (right) training. X1, X2 and X3 code the negative, middle and positive range in the X dimension (threshold=0.3). . . . .	138
A.38	The MI input maps with respect to visual input (in the Y dimension), before (left) and after (right) training. Y1, Y2 and Y3 code the negative, middle and positive range in the Y dimension (threshold=0.3). . . . .	139
A.39	The MI input maps with respect to visual input (in the Z dimension), before (left) and after (right) training. Z1, Z2 and Z3 code the negative, middle and positive range in the Z dimension (threshold=0.3). . . . .	140
B.1	Arm positions in different situations . . . . .	142
B.2	Muscle stretch map of PI after lesion . . . . .	143
B.3	Muscle stretch map of MI after lesion . . . . .	145

# Chapter 1

## Introduction

Artificial neural network models have become important computational tools in recent years. On the one hand, these models can be used to solve complicated practical problems that are difficult for conventional methods. Such problems include optimization problems [Hopfield & Tank, 1985; Ramanujam & Sadayappan, 1988; Angeniol *et al.*, 1988], control problems [Kuperstein, 1988; Bullock *et al.*, 1993; Mussa-Ivaldi *et al.*, 1991], pattern reorganization tasks [Gorman & Sejnowski, 1988; Cun *et al.*, 1989], etc. On the other hand, these models can also be used to study brain organization and disorders [von der Malsburg, 1973; Linsker, 1986; Pearson *et al.*, 1987; Grajski & Merzenich, 1990; Sutton *et al.*, 1994; Armentrout *et al.*, 1994; Weinrich *et al.*, 1994; Reggia *et al.*, 1996].

In this dissertation, a neural network model of motor control will be described, with simulated map formation both in sensory and motor cortex. The model approximates the closed-loop structure of mammalian motor control systems while remaining computationally tractable. It is based on a simplified arm that moves in 3D space. The arm has three pairs of antagonist muscles or muscle groups receiving motor control information and providing sensory information. Small portions of sensory and motor cortex corresponding to this arm are simulated. Training is done by supplying initial random stimulation to the motor cortex area and allowing the system to reach a stable state in response to each stimulus. After training, multiple cortical feature maps are measured and their characteristics and interrelationship are studied in order to understand more about motor control.

There are two motivations for the work described here. First, I wanted to determine whether a closed-loop, multi-layer motor control system could self-organize to form cortical feature maps that represent the characteristics of the simulated arm, and how these cortical features maps can be used to achieve consistent motor control. Motor control problems have long been of great interest to researchers in engineering, mathematics, computer science, and neuroscience [White & Sofge, 1992; Mel, 1990]. Controlling of arm position in 3D space has been studied intensively [Tarn *et al.*, 1991; Geffin & Furht, 1990; Nicosia *et al.*, 1989], mainly because of its application in robotic industries. Many modeling frameworks have been used to tackle this problem: kinematic versus dynamic, linear versus non-linear, feedback versus non-feedback, real-time versus trajectory planning, etc. Although great efforts have been made, this problem has not been solved satisfactorily with respect to efficiency, adaptability, robustness, etc. On the other hand, the problem of arm positioning is obviously solved successfully by mammalian animals, presumably based in part upon the feature maps existing in the cerebral cortex. However, the function of primary motor cortex (MI) in mammalian animals is currently not well understood. It is generally believed that MI makes use of feedback information via afferent sensory pathways to carry out

motor tasks. In particular, proprioceptive inputs play an important role in the formation of motor cortex outputs. Here proprioceptive inputs refer to sensory inputs from receptors inside muscles or tendons that reports the length and tension of muscles. Visual inputs are of course also important sensory feedbacks. How this kind of feedback information is processed and used by MI neurons is an important issue in identifying the function of MI. Computational neural network models can be trained to form feature maps used for motor control. Such a study will help us gain insights into the organization of primary motor cortex, and may generate new concepts and methods for use in automatic control systems.

The second motivation for this work was to investigate how multiple cortical feature maps simultaneously present in a region of sensorimotor cortex relate to each other. For example, in primary sensory cortex one can ask how the maps of muscle length (stretch) and muscle tension overlap or interrelate. In primary motor cortex, one can ask how both of these sensory maps relate to the motor output maps, and how maps of cortical activation of different muscles interact. In particular, I examined the following *temporal correlation hypothesis: when multiple feature maps exist in the same region of cortex, features that are temporally correlated will appear in map regions that are spatially correlated.*

The simulation results indicates that this closed loop system is capable of self-organizing during unsupervised learning. The motor output map appears to possess some properties seen in mammalian motor cortex, such as a distributed, multifocal representation of individual muscle groups. Thus, although this model is a substantial simplification of the corresponding biological system, it captures some fundamental principles underlying map formation in mammalian motor cortex. Also, these cortical feature maps were aligned in a way that reflect the mechanical constraints imposed by the model arm. For example, the sensory cortex map of the tension of a particular muscle group was found to align with the sensory cortex map of the length (stretch) of its antagonist muscle. In primary motor cortex, the output map of a muscle was found to align with the tension input map of the same muscle and the length input map of its antagonist muscle. With the presence of visual inputs, some post-training visual inputs maps were partially aligned with some motor output maps. Quantitative measurement indicated that the degree of alignment between two maps were monotonically related to the degree of temporal correlation between two features, thus verified the above hypothesis. We believe that such correlations are important for cortical motor neurons to consistently transform sensory input into motor output.

In summary, the primary contribution of this research are:

- Demonstrating that stable cortical feature map will form under unsupervised learning in a closed loop, multi-layered system.
- Showing that multiple cortical features maps in sensory and motor cortex align in a way reflecting temporal correlations due to the mechanical constraints of the model arm.
- Demonstrating that the alignment of visual input maps with motor output maps also reflect the temporal correlation between features during training process.
- Providing analysis of activation patterns in cortical feature maps and establishing the underlying principles of correlated and anti-correlated map features.

The rest of this dissertation is organized as follows. In Chapter 2, some background about map formation and motor control model will be given. Chapter 3 describes the structure of the motor

control model in detail. In Chapters 4 to 6, simulation results in three variations of motor control models are reported. Chapter 4 is about the model with proprioceptive input only. Chapter 5 uses visual input only instead of proprioceptive input. Chapter 6 gives the simulation results for the model with combined proprioceptive and visual input. In Chapter 7, some theoretical analysis concerning map formation and the interrelationships between map features is done using a simplified model. Chapter 8 summarizes the conclusions of this work and future directions for research.

## Chapter 2

# Background

This section describes past work on map formation in general, the principles of underlying map formation, and some previous motor control models using feature maps.

### 2.1 General Information on Self-Organizing Maps

Work on self-organizing maps, both in computer science and in computational neuroscience, was initially motivated by the observation that such maps are widely observed in mammalian nervous systems [Penfield & Rasmussen, 1950; Hubel & Wiesel, 1962]. The term *maps* here refers to the order-preserving representations in the cortices of mammalian systems of the outside sensory or motor control space. There are mainly two classes of maps: topographic maps and feature maps (the latter also being called computational maps). The term *topographic map* refers to the fact that the sensory surface or motor control space is represented in cortices in topographic order. The term *feature map* means neurons in sensory or motor cortices respond to certain features of the sensory or motor space.

For topographic maps, the similarity of input patterns is measured in terms of geometric proximity of the input patterns. Therefore the cortex is a direct reflection of the spatial ordering of the outside world it represents. For example, for primary somatosensory cortex (abbreviated SI), there is a representation of the skin surface across the cortex [Freeman, 1979]. Every region of the body surface has a corresponding area in SI, and adjacent regions of the body generally have adjacent corresponding areas in SI. For primary motor cortex (called MI), there is a similar topologically preserved mapping from muscles of the body to the cortex (Fig. 2.1), although this is combined with feature maps at the detailed level, and is more controversial [Donoghue *et al.*, 1992].

On the other hand, for feature maps the similarity of input patterns is measured in terms of functional similarity of input patterns, for any particular function. Primary visual cortex (called VI) is an example of a feature map [Hubel & Wiesel, 1959; Hubel & Wiesel, 1962]. Many neurons in VI respond maximally to input stimuli (such as lines) that have particular orientations. Therefore a transformation is needed to change the information of size and position of input signals into orientation information. Input patterns with similar orientations, rather than similar location, are said to be similar. Many of the details of the mechanisms of this transformation of information in the brain are still unknown [Hubel & Wiesel, 1962; Weyand *et al.*, 1986; Chapman *et al.*, 1991].



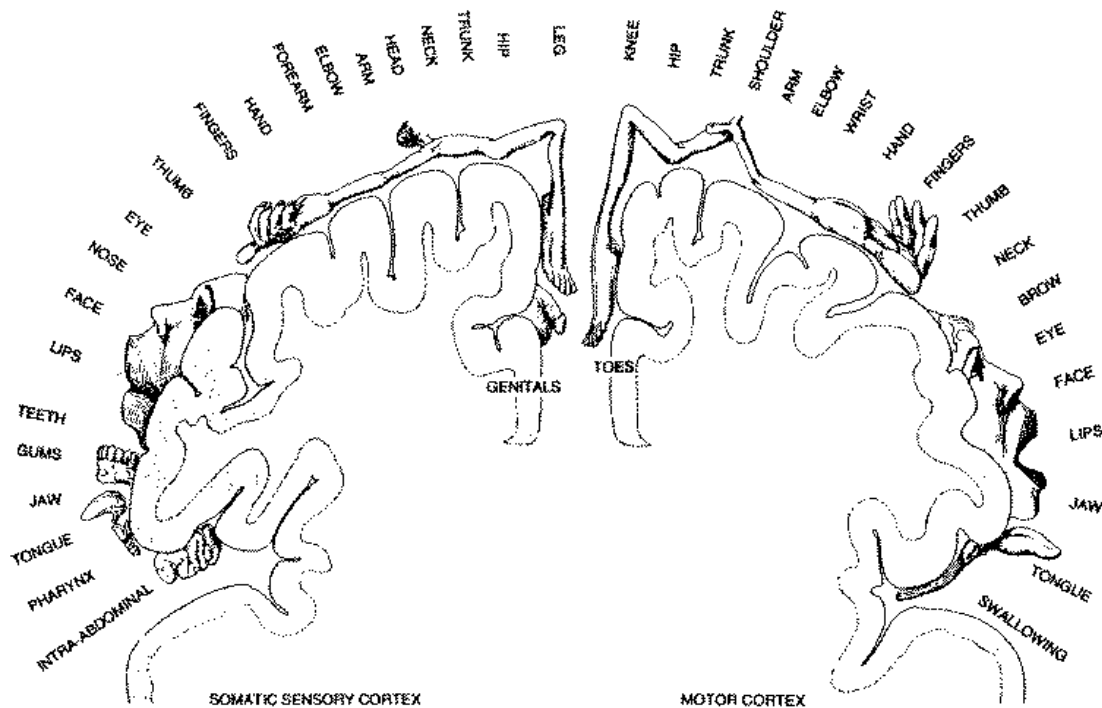


Figure 2.1: Topographic maps in human somatosensory and motor cortex. Each part of the human body is represented by corresponding area in sensory and motor cortex in a topographic preserved fashion. Stimulating a particular body surface area will activate corresponding somatosensory cortex region. And activation of a motor cortex region will cause muscle contraction in the corresponding area. (Picture taken from “The Brain”, A Scientific American Book), 1979

### 2.1.1 Using Maps for Computational Purposes

Map formation can be used for computational purposes. Von der Malsburg did some of the first work on simulating map formation [von der Malsburg, 1973]. Kohonen proposed a computational model which can be used for feature map map formation [Kohonen, 1982; Kohonen, 1995]. Although it is biologically unrealistic, it serves as a way to organize information.

In Kohonen’s model, there is a network consisting of a simple input layer and a simple output layer, which are fully connected. Usually both layers are two-dimensional. The input layer receives input patterns, which are viewed as being ordered according to some definition. The output layer has lateral connections and is ordered naturally by means of the relationship of neighboring nodes. The purpose of these lateral connections is to insure that similar input patterns will generate similar responses in the output layer. For this to occur, a measurement of similarity of input patterns must be defined. Usually, the similarity of input patterns, which are represented as vectors, is measured by the inner product [Kohonen, 1989].

The activation rule of Kohonen’s model is based on the inner product of an input pattern and the incoming weight vector:  $a_i = \sum_{j=1}^n w_{ij}in_j$ . Here  $a_i$  is the activation of output node  $i$ ,  $w_{ij}$  is the weight connecting input node  $j$  and the output node  $i$ ,  $in_j$  is the  $j$ ’th component of the input

pattern. According to the activation rule, the node in the output layer which has an incoming weight vector most similar to the input pattern tends to get the highest activation (although some normalization procedures may result in distortion [Sutton & Reggia, 1994]). The learning rule is such that when a node matches an input pattern, its incoming weight vector is adjusted to be more similar to the input pattern, so that next time that same input pattern occurs this node is more likely to respond. For the winning node  $c$  (i.e., the most highly activated node), its learning rule is:  $\Delta w_{cj} = \eta(in_j - w_{cj})$  for all  $j = 1 \dots n$ . Here  $\eta$  is a constant controlling learning speed.

Besides the winning node  $c$ , its neighboring nodes are also allowed to learn from this input pattern, so that neighboring nodes will develop similar weight vectors as node  $c$ , and therefore have similar responses to input patterns. The neighborhood is defined to be the region within which nodes can send activation to each other via excitatory lateral connections. The neighboring nodes use the same learning rule as node  $c$ . In Kohonen's model, the initial size of the neighborhood is defined to be the entire output layer. Therefore all the incoming weight vectors are adjusted to become similar and the map is conceptually compressed in the center of the region. By gradually decreasing the neighborhood region, the output map can be expanded smoothly. Kohonen's model can be used to represent information in an efficient way so that similar input information will activate nodes adjacent each other. This simple map can also be used to represent information at different abstract levels.

Kohonen's model on map formation has been used in many computational applications. Some examples: to develop a speech recognition device that can recognize a large vocabulary of isolated words by their trajectories across a phonetic map surface [Kohonen, 1987], to solve optimization problems such as the traveling salesman problem (TSP) [Angeniol *et al.*, 1988], and for the motor control problem described above [Kuperstein, 1988; Ritter *et al.*, 1989; Walter & Schulten, 1993].

### 2.1.2 Biological Modeling Using Maps

Computational models of map formation not only can be useful computational tools, they also can be used to simulate biological systems. One example is von der Malsburg's computational model in simulating the orientation activated neurons in mammalian primary visual cortex [von der Malsburg, 1973]. The occurrence of orientation sensitive cells in the primary visual cortex makes it different from other area in cortex. The mechanism underlying such an organization has been investigated by many scientists since their discovery [Hubel & Wiesel, 1963]. Von der Malsburg described the first computational model of primary visual cortex showing the self-organization of orientation sensitive cells via learning [von der Malsburg, 1973]. Other models of visual cortex, such as those modeling the ocular dominance columns in VI, have also been developed [Miller *et al.*, 1989; Tanaka, 1991]. These models used both mathematical analysis and computer simulations to analyze conditions under which the ocular dominance columns occur, and to explain the resultant ocular dominance columns patterns.

In Kohonen's model as described above, the winning node is picked globally, and the neighborhood size changes dramatically during the training process. Moreover, the input layer is fully connected to the output layer. All of these characteristics are biologically implausible. Thus, in many biological modeling applications, researchers have adopted different architectures. These models usually have limited connections between input and output layers and restricted areas of lateral connections. One of these models was motivated by the observation of feature map formation in the mammalian primary visual cortex before any visual experiences [Linsker, 1986]. This

model is a multi-layer feed forward network, with overlapped forward projections and limited lateral connectivities. Linsker demonstrated that, with a simple Hebb-type learning rule, a network can self-organize to form orientation selective columns based on purely random activation at the input layer. This is different from the models previously discussed in this section, which use co-related activation patterns as inputs. It means that the necessary connections and some of the underlying updating principles, instead of the feature inputs, are sufficient to account for the map formation in the cortex.

Pearson *et al.* proposed a model of topographic map formation that avoided many of the limitations of Kohonen’s model [Pearson *et al.*, 1987]. In this model, each input node was connected to its corresponding node in the output layer and its surrounding nodes, forming a coarse topographic map at the beginning of training; the receptive fields of the output nodes overlapped extensively. The nodes in the output layer had local internal connections. The training was done by supplying patched activation patterns in the input layer. During training, this coarse topographic map was refined and the receptive fields became smaller and more concentrated. This model is biologically plausible not only in the sense of local interactions of neurons, but also in that it shows effective reorganization after a change in input patterns.

Grajski and Merzenich proposed another model simulating map formation in somatosensory cortex [Grajski & Merzenich, 1990]. This model is similar to Pearson’s model in term of connectivity and the training method, except an intermediate layer was added to the network to simulate subcortical neurons. This layer is used to increase the area of projection from the input layer to the cortical layer and to allow the subcortical layer to dynamically affect the cortical inputs. Grajski’s model has shown refinement of the initially coarse topographic map during training as well as the map reorganization due to repetitive stimulation and lesioning. This model is an improvement when compared with Pearson’s model in that it maintains the inverse magnification rule during map reorganization,<sup>1</sup> and is therefore more plausible in simulating map formation in mammalian cortex.

A model using competitive distribution of activation to simulate topographic map formation in somatosensory cortex (referred to as the SI model in subsequent discussion) has also been developed. The structure of this model is quite similar to the one used in [Pearson *et al.*, 1987]. In the SI model, each input node is connected to its corresponding output node and that node’s neighbors within a certain distance, forming an initial coarse topographic map from the input layer to output layer. The coarse topographic map at the beginning can be regarded as genetically predetermined. The training pattern is a hexagonal activation patch of radius of one or two, uniformly distributed over the entire input layer. It has been shown that, after training, the coarse topographic map becomes refined. The receptive fields become more regular; and the incoming weight vectors becomes Gaussian shape in distribution [Sutton *et al.*, 1994; Armentrout *et al.*, 1994].

The competitive distribution of activation can also be used in models forming feature maps. One such models has a structure similar to von der Malsburg’s model, and is trained with the same set of input patterns [Weinrich *et al.*, 1994]. The training of this model produced results similar to those in von der Malsburg’s model, such as formation of clusters, and decrease of nodes tuned to multiple orientations. The only difference observed was that the competitive activation model produced smaller, more activated clusters than von der Malsburg’s model.

---

<sup>1</sup>The inverse magnification rule, which is observed in biological experiments [Jenkins *et al.*, 1990], states that there is an inverse relationship between cortical magnification and receptive field size.

Models using competitive distribution of activation have demonstrated the ability to reorganize after focal map damage. It has been shown that the above SI model exhibits spontaneous map reorganization in response to a cortical lesion [Sutton *et al.*, 1994], unlike some earlier models [Grajski & Merzenich, 1990]. The model exhibits a two-phase reorganization process. Immediately after the lesion, the receptive fields of cortical nodes adjoining the lesioned area shift towards the lesioned area. This shift is caused by the dynamic redistribution of activation due to the competitive distribution of activation. After continued training following a lesion, more of the finger region initially represented by the lesioned region is now represented by nodes in the surrounding region. This second-phase change is caused by the shift of weight vectors and is triggered by the first-phase of map reorganization [Sutton *et al.*, 1994].

## 2.2 Cortical Lateral Inhibition

A stimulus to cortex via sensory pathways has the characteristics of an excitatory area surrounded by an inhibitory region. This phenomenon can be observed in somatosensory cortex as well as visual cortex [Mountcastle, 1978]. It also can be produced by direct activation of a small region in neocortex [Hess *et al.*, 1975; Gilbert, 1985]. The activation pattern in the primate cortex is as follows:

- A central excitatory region with radius of 50 to 100  $\mu\text{m}$ .
- An inhibitory region surrounding the central excitatory region reaching up to a radius of 200 to 500  $\mu\text{m}$ .

A weaker excitatory region surrounding the inhibitory penumbra reaching up to a radius of several centimeters may also be observed. This activation pattern is usually referred to as a “Mexican Hat” pattern (Fig. 2.2). This phenomenon is usually attributed to the lateral interactions between cortical neurons.

In the traditional view, the inhibitory part of the “Mexican Hat” activation pattern is attributed to direct lateral inhibitory synaptic connections. According to this view, when a cortical site is activated, it suppresses activation of nearby cortex because of its direct or indirect lateral inhibitory synaptic connections to nearby cortex (Fig. 2.3). Accordingly, most past computational models of map formation have used this kind of mechanism to produce central excitatory, peristimulus inhibitory activation pattern [Kohonen, 1989; Pearson *et al.*, 1987; von der Malsburg, 1973]. However, the circuitry in neocortex is poorly understood at present. Some of the observations are difficult to reconcile with the inhibitory lateral connection scheme. For example, most inhibitory links in cortex are vertical or intracolumnal rather than lateral. The lateral inhibitory connections are sparse and appear to be mismatched to the distribution of peristimulus inhibition. Such facts have led to the investigation of alternative mechanisms underlying the “Mexican Hat” activation pattern, such as competitive distribution of activation [Reggia *et al.*, 1992].

Models using competitive distribution of activation have been shown to be quite useful in modeling neurophysiological phenomena. Competitive distribution of activation has been successfully used for feature map formation as well as topographic map formation [Sutton *et al.*, 1994; Armentrout *et al.*, 1994; Cho & Reggia, 1994; Weinrich *et al.*, 1994]. These models provide for peristimulus inhibition without lateral inhibitory connections. The main idea of competitive distribution of activation is that the spread of activation is based on not only the connection strength

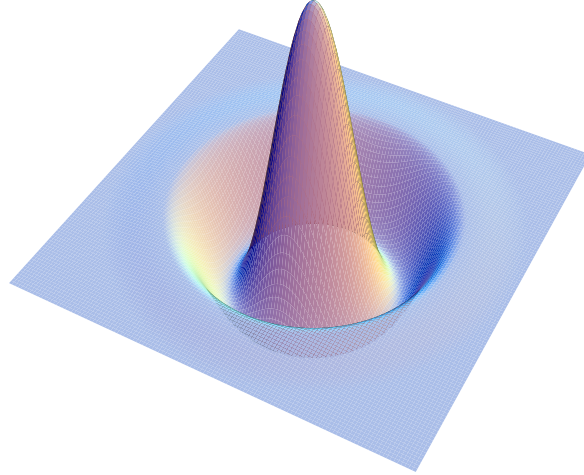


Figure 2.2: “Mexican Hat” like activation pattern: the two dimensional x-y plane represents the extension of cortical surface, the height (z axis) represents the level of activation.

to the destination neurons, but also the activation level of those destination neurons. Destination neurons with higher activation levels tend to receive more activation than the ones with lower activation levels. Therefore, populations of neurons (e.g., cortical columns) with different activations tend to “compete” for activation. Formally, this mechanism is implemented using a formula such as:

$$out_{ji}(t) = \frac{a_j(t)w_{ji}}{\sum_k a_k(t)w_{ki}} \cdot a_i(t) \quad (2.1)$$

where  $out_{ji}(t)$  is the activation that node  $j$  receives from node  $i$ , and  $k$  ranges over all the nodes to which node  $i$  sends activation (Fig. 2.4). In this formula, the activation that node  $j$  receives from node  $i$  is decided not only by the connection strength  $w_{ji}$ , but also by the activation of node  $j$  as well as other nodes to which node  $i$  send connections. If node  $j$  has a higher activation level than node  $k$ , then it tends to gain more input from node  $i$ . Since total activation from node  $i$  is fixed at time  $t$ , other nodes  $k$  will receive less input activation from node  $j$ . This kind of inhibition is sometimes called *virtual inhibition*. Competitive activation was used in several of the models of map formation described above, and generally has produced results that are qualitatively the same as lateral inhibitory links when simulating map formation.

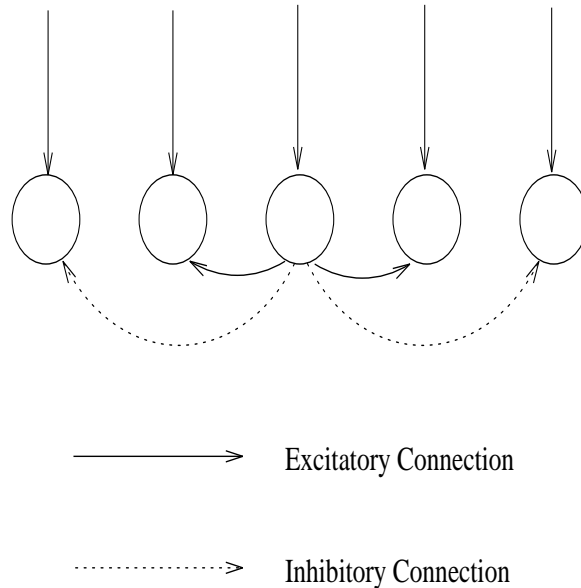


Figure 2.3: A schematic diagram of lateral connections to produce central excitatory, peristimulus inhibitory activation patterns. Other implementation may have the same effect.

## 2.3 Feature Maps for Motor Control

In this section, some past computational models of motor control using feature maps are discussed. First, a class of models of motor control developed by others using visual input information is described. Second, a feature map model of proprioceptive sensory afferents developed at the University of Maryland is introduced. This latter model, the first computational model of proprioceptive cortex<sup>2</sup>, is the starting point for my efforts to develop a motor control model incorporating proprioceptive sensory information.

### 2.3.1 Modeling of Motor Cortex: Visuo-motor-coordination

Several computational models of motor cortex have focused on feature map formation for visuo-motor coordination [Kuperstein, 1988; Walter & Schulten, 1993; Ritter *et al.*, 1989], mainly because of its application potential in industrial robotics. Visual input is typically fed to a single layer network which directly produces motor output. The motor output directs the movement of a robot arm and thus changes visual input. Through training, visuo-motor coordination is achieved. Other motor control models which invoke a more complex architecture or more biologically plausible ingredients were also proposed [Mel, 1988; Mel, 1990; Burnod *et al.*, 1992].

Ritter *et al.* proposed an algorithm of visuo-motor coordination to control a robotic arm [Ritter *et al.*, 1989]. In this model, visual input is obtained from two cameras. The input pattern is two pairs of coordinates, reflecting the hand position observed by the cameras. The output is composed of joint angles and other relevant elements. Between the input layer and output layer, there is a neural network layer, connected like the one in Kohonen's model, doing the information

---

<sup>2</sup>As noted earlier, proprioception refers to sensory input concerning muscle tension, muscle length, joint position, etc.

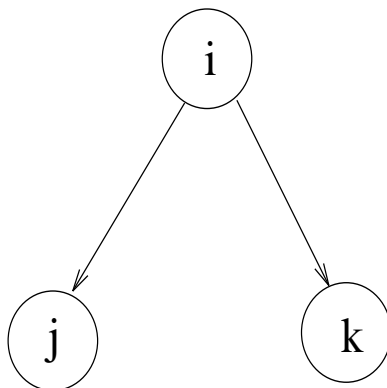


Figure 2.4: Competitive activation mechanism: node  $i$  sends activation to nodes  $j$  and other nodes  $k$  (maybe more than one). The distribution of activation is based not only on the weight from nodes  $i$  to nodes  $j$  and  $k$ , but also on the activation level of nodes  $j$  and nodes  $k$ .

processing. For each node in this layer, there is an incoming weight vector from the input layer, and an outgoing weight vector to the output layer. The incoming vector is adjusted towards the input pattern. The outgoing vector is adjusted towards an estimated position generated from the movement error viewed by two cameras. The learning algorithm is thus called an “extended self-organizing feature map algorithm”. This algorithm is similar to the Kohonon’s “self-organizing feature map algorithm”, except that, instead of only winner node updates weights, the learning rule in this model is set such that in each learning step, every node learns to adjust its incoming weight vector toward the input pattern, but the winning node learns most, and increasingly distant nodes learn progressively less. The model is trained to learn arm kinematics by finding the appropriate joint angles for each target location, and to learn arm dynamics by generating appropriate joint torques necessary to accelerate the end effector of the robot arm from a given position to a specified velocity. It was shown that both arm kinematics and arm dynamics can be learned in this model.

Walter and Schulten reported the implementation of two learning algorithms for visuo-motor control of an industrial robot (PUMA 562) [Walter & Schulten, 1993]. Their system also has two cameras providing visual information to the neural network, and a multi-joint arm controlled by the output commands of the neural network. The task learned by the robot is to position its end effector at certain positions in the space. The first learning algorithm they used is the “extended self-organizing feature map algorithm”, described above. The second algorithm, the “neural gas” algorithm [Martinetz & Schulten, 1991], is similar to the first in the sense that all the nodes are allowed to learn for each learning step, with different amounts. However, in the “neural gas” algorithm, the network layer has no topologic relations. There are no lateral connections in this layer. The degree of neighborhood is not defined by the geometric distance, but dynamically by the similarity of the incoming weight vectors to the input vectors. The node with incoming weight vector most similar to the input vector will learn most. Other nodes learn less, based on their ranks in the ordered sequence of similarity, with an exponentially decreased amount of adjustment. The “neural gas” algorithm defines the neighborhood relationship dynamically. Thus it allows a more flexible topology, especially when the spatial relationship is unknown or inhomogeneous. It has been shown that both algorithms can be used in visuo-motor control of an industrial robot [Walter & Schulten, 1993], and that a topology preserving map forms after training. With only

3000 learning steps, the system is able to position its end effector with a precision of 0.1% of the linear dimension of work space.

Kuperstein proposed another model using a self-organized neural network to achieve hand-eye coordination [Kuperstein, 1988]. The task performed is to position a simulated robot arm at a certain location in space, and with a certain orientation. This model is more complicated than the previously discussed models doing similar tasks. The robot not only has camera-like eyes providing visual input, but also can adjust the orientation of the eyes with different tensions of eye muscles. Therefore, in order to obtain a correct assessment of spatial locations, the visual information as well as the activations of eye muscles are necessary inputs to the network. The training of the network is as follows. Self-produced motor signals are first generated to move the arm to certain locations in space. The images captured by the two “eyes” are combined to produce visual maps. The muscle activations of the two “eyes” are also combined to produce gaze maps. The visual maps and gaze maps are then combined to compute the necessary motor signal. These computered motor signals are then compared with the initially generated motor signals. The learning rule is set such that the differences between the initially random motor signals and the computed motor signals from the visual feedback information are minimized. Computer simulations showed that, after training, the model is capable of performing the task with an average position error of 4% of the arm’s length and with an average orientation error of 4°.

Another motor control system which involves multiple sensory and motor control layers was proposed and implemented by Mel [Mel, 1988; Mel, 1990]. This system, called robot Murphy, was designed to control a robot arm to grab an object in a 2D plane, with a visual input provided by a color video camera. The control part of Murphy consists of four layers of neuron-like units: visual-field population, hand-velocity population, joint-angle population and joint-velocity population. The video camera provides Murphy with visual information (hand positions, target positions, obstacles etc.) and hand-velocity information. The joint-angle and joint-velocity layers were used to control the motors in the robot arm joints. All four layers were initially coarse coded, with Gaussian shape receptive and/or projective fields. A learning algorithm, Sigma-Pi learning, was used to avoid introducing non-linear intermediate units. Murphy was trained to learn both forward kinematics from the joint-angle layer to visual-field layer and the inverse kinematics from hand-velocity layer to joint-velocity layer. The redundant control dimensions of robot arm enable Murphy to plan multiple routes to a target and could successfully perform target reaching tasks while avoiding obstacles.

The above motor control models have emphasized reaching tasks, although some biological structures were borrowed to achieve this goal. On the other hand, recent developments in neurophysiology have motivated the building of computational models that can be used to explain biological results. One recent discovery about the motor cortex is its use of directionally tuned neurons and population coding [Georgopoulos *et al.*, 1986]. Some computational models using directionally tuned units to achieve motor control tasks have also been developed [Burnod *et al.*, 1992]. These models emphasize the role that directionally tuned neurons play in the motor cortex, although motor control tasks can also be performed. In [Burnod *et al.*, 1992], a motor control model was described. This model combines visual and somatic inputs in the sensory cortex layer and generates motor commands in the motor cortex layer. In this model, each processing unit in a layer corresponds to a cortical column rather than to an individual neuron. Each layer can be further divided into sublayers. In the input layer, visual and somatic input is received in different sublayers. Simple Hebbian learning was applied for both intralayer and interlayer connection learning, while



spontaneous movements were used during learning. After training, the motor units were tuned to preferred directions which rotate with initial arm position, while the population vectors were parallel to trajectories. This is consistent with neurophysiological results.

These previous neural network motor control models, although having exhibited substantial success in fulfilling the goals of investigators, have some limitations, especially when viewed from a biological perspective. In most models, motor and visual sensory functions are combined into one network layer, and there is usually no proprioceptive feedback. These facts plus the unrealistic connectivity are biological implausible. These models do not tell us much about the maps inside the brain, especially the primary motor cortex. The model proposed in [Burnod *et al.*, 1992] is more biologically realistic in structure. It also includes proprioceptive input. However, the proprioceptive information was given in the form of a 3D vector representing arm position. No realistic proprioceptive afferents were modeled as in the research described here. Therefore this model still lacks plausibility in proprioceptive feature map formation, and in how the sensory feature map affects the feature map in motor cortex and motor behavior. The model that I propose here is different from these previous motor control models in that it accounts for the biological feasibility as well as the motor control tasks. This model tries to simulate the structure and behavior of the mammalian sensory motor system. In this model, realistic (although simplified) sensory and motor control feature maps are simulated and their relationship is investigated. Thus the model serves as a tool to study the cortical activity of biological systems, as well as exploring a new approach to motor control. In addition, we use the model to examine how multiple simultaneously present feature maps align with each other, something which has not been done previously. Based on the fact that biological motor control systems are superior to any artificial motor control system, I believe that the improvement of artificial systems may be possible when deeper knowledge of mammalian cortex is obtained.

### 2.3.2 Proprioceptive Cortex Map

A model simulating map formation in primary sensory cortex based on proprioceptive input from an arm has been implemented in our research group [Cho *et al.*, 1993]. This model (referred to as the PI model) is motivated by the fact that proprioceptive input plays an important role in map formation in the brain and in the coordination of motor control [Asanuma, 1989]. The motor control model described in this dissertation incorporates the basic concepts of the PI model as one component.

In this model, there is a simulated model arm that moves in a three dimensional space. Six generic muscles, controlling the movement of the arm, also provide proprioceptive information about the length and tension of each muscle. The details of the arm model will be described in the next chapter.

The structure of the network is illustrated in Fig. 2.5. Like most previous models of map formation, there is a single input layer and a single output layer. The length and tension of the six muscles serves as the proprioceptive input in this network. The proprioceptive cortex layer (also referred to as PI layer) consists of 400 nodes, with hexagonal tessellation. Each cortical node is connected to its 6 neighbors. The connection from proprioceptive input layer to cortical layer is fully connected.

The methods used in this network are competitive distribution of activation along with competitive learning, which will be described in detail later. The weights from arm proprioceptive

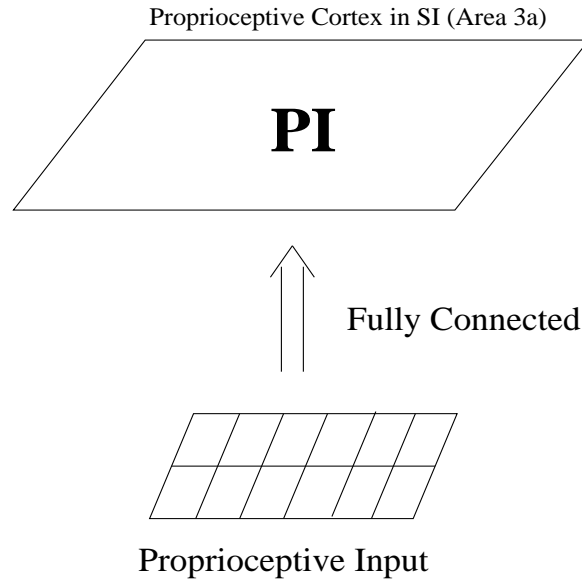


Figure 2.5: Network architecture: the input layer has 12 nodes representing lengths and tensions of 6 muscles; The cortical layer has 400 nodes with lateral connection of radius 1. The input layer and the cortical layer are fully connected.

input layer to cortical layer were initialized randomly, forming a poorly defined map. Studies have focused on examining map formation in PI after being trained with proprioceptive input based on the random movement of the arm [Cho & Reggia, 1994]. The training procedure in each learning cycle is:

- generate six random values representing muscle activations to set the arm position;
- compute the proprioceptive input based on the arm position;
- send activation from sensory neurons to the cortical layer; and
- train the network connections from proprioceptive input layer to cortical layer, using competitive learning method.

After training, the following results were obtained:

- Most proprioceptive cortex nodes were tuned to the length or tension of a particular muscle. Nodes tuned to the same muscle length or tension tended to group together to form clusters. The size of clusters became more uniformed after training. Moreover, the group of nodes tuned to the lengths of antagonist muscles tended to push apart from each other, reflecting the mechanical constraints imposed by the movement of the arm (antagonist muscles can not be stretched simultaneously, and thus only one tends to be active at any time).
- Among the nodes which were tuned to multiple inputs, the number of nodes which were tuned to implausible input pairs decreased to zero and the number of nodes tuned to plausible input pairs increased significantly as the result of training.

- A spatial map of hand positions was also formed.

The above results show that, after training, feature maps formed in proprioceptive cortex layer. These maps caught some characteristics of the model arm implicitly contained in the input patterns. Varying model details resulted in variations of map details in predictable ways.

This proprioceptive model provided us with an initial understanding of proprioceptive feature maps, and thus it represents preliminary work for the research described in this dissertation. However, the PI model is limited in the sense that it has no motor cortex involved. Therefore, all the motor output (or muscle activation) is hypothetical or artificial. Also it does not have a closed-loop architecture in which the sensory feedback could alter motor output. The PI model did provide a starting point from which to build a more complicated and biologically plausible motor control model.

## Chapter 3

# The Motor Control Model

This chapter describes in detail the motor control model used in this research, including the arm model, network structure, activation and learning rules, and training method. The methods used to measure the resultant cortical feature maps are also described.

This motor control model simulates the closed-loop path of information flow in the nervous system. In the first stage of the simulations, only proprioceptive input is used as sensory feedbacks. Later, visual inputs (or combined proprioceptive inputs and visual inputs) are used as sensory feedback. Fig. 3.1 gives out a schematic diagram of the closed-loop motor system involving just proprioceptive input. The activations of muscles direct arm movements. Proprioceptive information about the muscles is then fed into the primary sensory cortex, which supplies this information to primary motor cortex, thus influencing the motor output. The closed-loop system has the advantage of “knowing” the results of certain motor commands from sensory feedback and adjusting motor output based on such feedback.

### 3.1 The Arm Model

The model arm simulated here is a significant simplification of biological reality. It is not a neural model. The model arm has an upper arm and a lower arm, connected by the elbow, that moves in a three dimensional space. There are six generic arm muscles or muscles groups, with one pair of muscles groups (extensor and flexor) controlling the movement of the lower arm, and two pairs of muscles groups (extensor and flexor, abductor and adductor) controlling the movement of the upper arm (Fig. 3.2).

For a particular set of activation values of agonist and antagonist muscles, the corresponding joint is positioned at a particular angle. Therefore the length of each muscle is determined. Biologically, this kind of length information is measured by the receptors in muscle spindles embedded in parallel with muscle fibers. The tension information of muscles is measured by receptors in Golgi tendon organs. The length and tension information of muscles is then transmitted to proprioceptive cortex in SI (Brodmann area 3a). I will refer this primary proprioceptive cortex area as PI in later discussions.

Fig. 3.3 shows a generic joint where appendage  $OP$  of length  $l$  is moved  $180^\circ$  around the axis which is perpendicular to the plane  $AOP$  and passes through the origin  $O$ . Thus, movements of the endpoint of the appendage,  $Q$ , define a semicircle  $APB$ . The movement of appendage  $OP$  is controlled by changing the lengths of the muscles  $l_1(XZ)$  and  $l_2(YZ)$ . Both muscles are attached to the mid-point of the appendage,  $Z$  (i.e.,  $\overline{OZ} = l/2$ ) on one hand. Also muscle  $XZ$  is attached

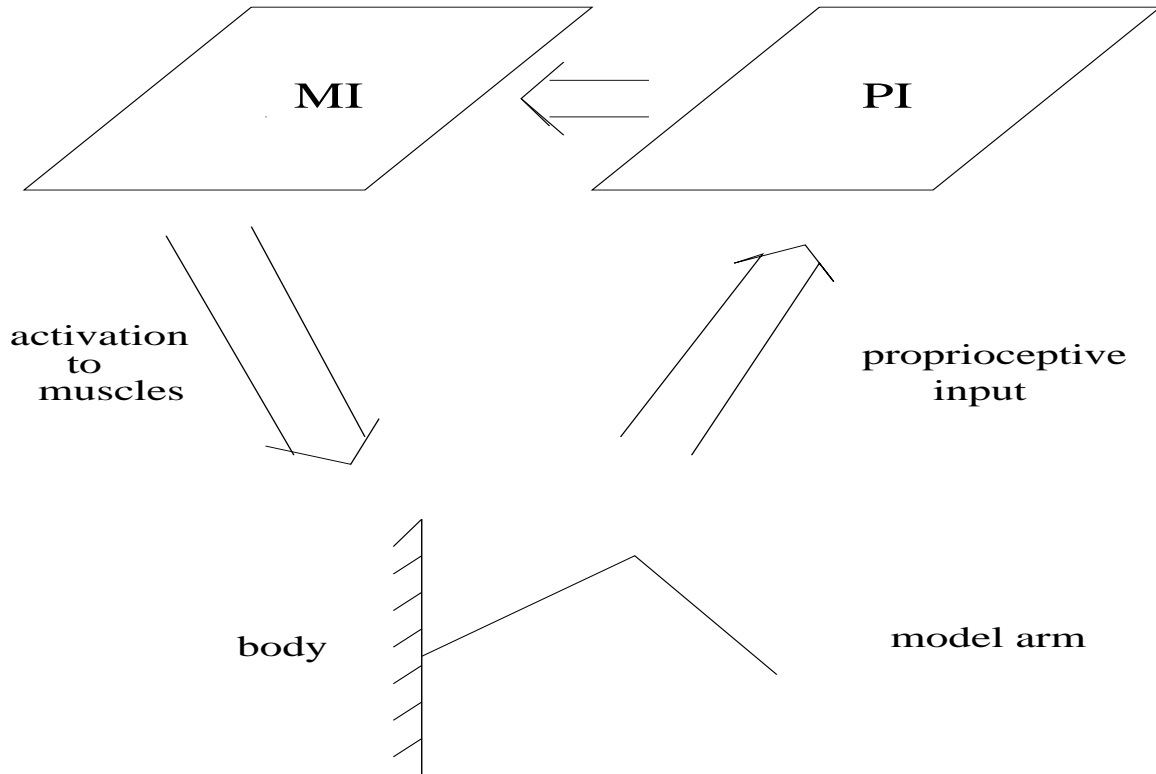


Figure 3.1: Schematic diagram of the closed loop motor system: the model arm, directed by motor neuron activity, supplies proprioceptive information to proprioceptive cortex (PI). This proprioceptive information then influences neuron activities in the primary motor cortex, therefore changing the motor output commands.

to point X and muscle YZ to point Y, respectively, which are located distance  $l/2$  apart from the origin on opposite sides (i.e.,  $\overline{OX} = \overline{OY} = l/2$ ).

For convenience, in our model the resting position of an appendage is defined as perpendicular to the axis to which the pair of muscles are attached (i.e.,  $OP$ ). Joint angle  $\theta$  denotes the angle between this resting position ( $OP$ ) and the current position of the limb ( $OQ$ ). We define the joint angle as a function of difference between the input activation level of agonist and antagonist muscles which control the joint.

$$\theta = \frac{\pi}{2}(f(in_{ag}) - f(in_{ant})) \quad (3.1)$$

where  $\theta$  ranges from  $[-\frac{\pi}{2}, \frac{\pi}{2}]$ ;  $f(in_{ag})$  and  $f(in_{ant})$  are the functions of activation of agonist and antagonist muscles, representing any of the three pairs of muscles. Function  $f$  maps muscle activations into values ranging from 0 to 1 so that appropriate joint angles can be calculated. In the model described in this dissertation,  $f$  is the identity function because the network parameter setting has ensured the activation of muscle activations to be within the range of 0 to 1.

The length of a muscle can be easily derived from the joint angle, according to Fig. 3.3.

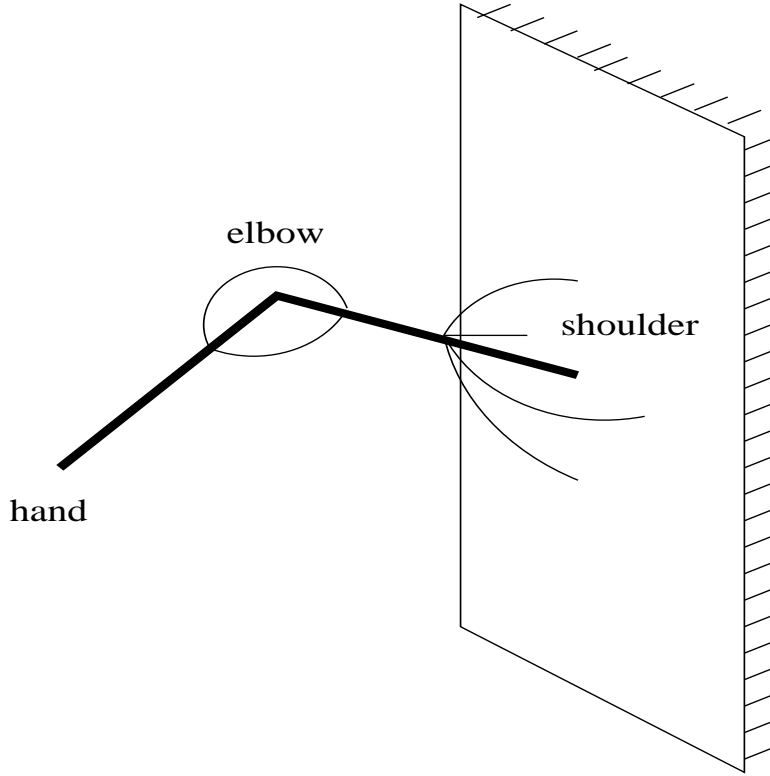


Figure 3.2: Model arm. Three pairs of muscles (indicated by curves) control the movement of upper arm and lower arm (indicated by bold line segments). Two pairs of antagonist muscles control the upper arm, while the third pair of antagonist muscles controls the lower arm.

$$l_1 = l \cos \frac{1}{2}(\frac{\pi}{2} - \theta) \quad (3.2)$$

$$l_2 = l \sin \frac{1}{2}(\frac{\pi}{2} - \theta) \quad (3.3)$$

where  $l$  is a constant representing length of an appendage.

To see this, consider  $\triangle OYZ$ , an isosceles triangle with  $\overline{OY} = \overline{OZ} = l/2$ . Let  $W$  be on  $YZ$  such that  $OW \perp YZ$ , so  $\triangle OWY$  is a right triangle with

$$\angle YOW = \frac{1}{2}(\frac{\pi}{2} - \theta) \quad (3.4)$$

and

$$\overline{YW} = \frac{l_2}{2}. \quad (3.5)$$

From Eqs. 3.4 and 3.5, we get

$$\sin \frac{1}{2}(\frac{\pi}{2} - \theta) = \frac{\overline{YW}}{\overline{OY}} = \frac{l_2/2}{l/2} = l_2/l,$$

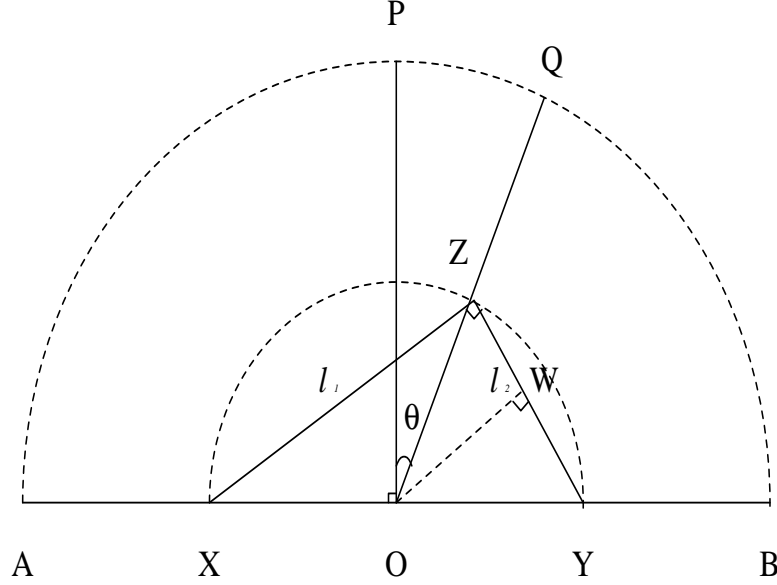


Figure 3.3: Generic joint

thus, we have Eq. 3.3.

Now consider  $\triangle XZY$ . Since point  $Z$  is on a semi-circle with center  $O$  and diameter  $l$ ,  $\angle XZY = \frac{\pi}{2}$ . Thus, we have

$$\begin{aligned} \overline{XZ}^2 + \overline{YZ}^2 &= \overline{XY}^2 \\ l_1^2 + l_2^2 &= l^2 \\ l_1 &= \sqrt{l^2 - l_2^2}. \end{aligned}$$

Substituting Eq. 3.3 for  $l_2$ , we have Eq. 3.2 since  $\frac{1}{2}(\frac{\pi}{2} - \theta) \in [0, \frac{\pi}{2}]$ .

The tension of each muscle is measured by the Golgi tendon organs that are arranged in series with muscle fibers. These receptors respond strongly when the muscle actively contracts. Passive stretching of the muscle also activates the Golgi tendon organ but not as much [Kandel & Schwartz, 1985]. The tension of each muscle is decided jointly by motor neuron activation as well as the length of muscle:

$$T_{ag} = f(in_{ag}) + T \cdot l_{ag} \quad (3.6)$$

$$T_{ant} = f(in_{ant}) + T \cdot l_{ant} \quad (3.7)$$

where  $T$  is the passive tension constant.  $T$  is usual small so that the first portion  $f(in_i)$  (activation tension) is much stronger than the second portion  $T \cdot l_i$  (passive tension).

## 3.2 Network Structure

Fig. 3.4 shows the structure of the interconnected neural networks in the motor control model. There are four layers of neural elements: proprioceptive input layer, proprioceptive cortex layer,

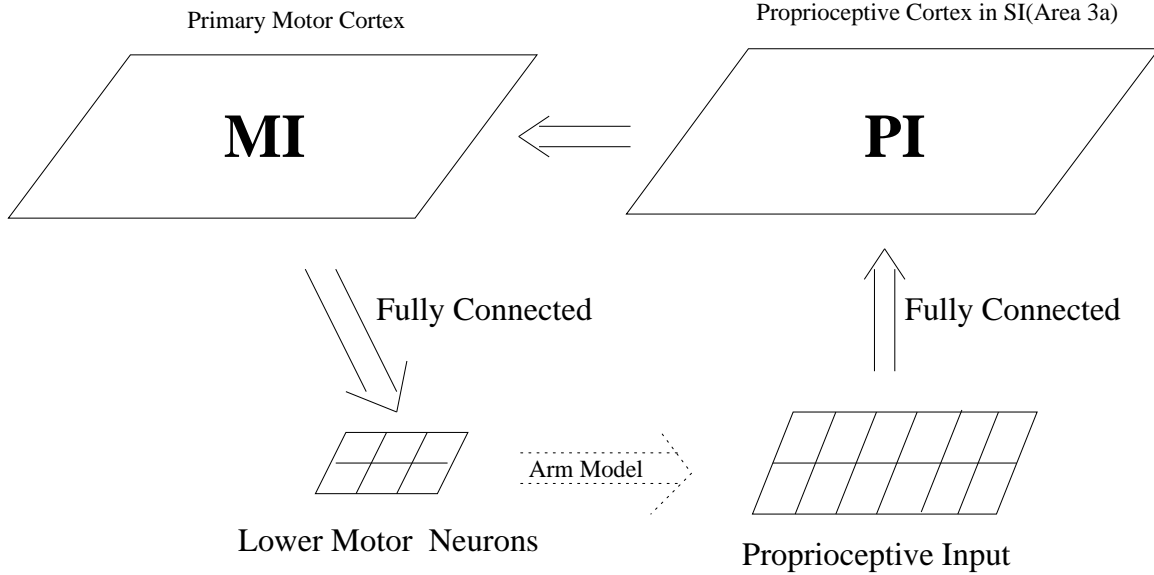


Figure 3.4: Network architecture of the motor control model: twelve proprioceptive receptor elements form the proprioceptive input layer and are fully connected to the PI layer. The proprioceptive cortex layer PI and primary motor cortex layer MI are two dimensional arrays of elements with lateral connections. The projection from PI to MI is partial, with a coarse topographic ordering. Each MI element is connected to the six lower motor neuron elements. The transformation of activity in lower motor neurons to proprioceptive input is done by a simulated arm represented by Equations 3.1, 3.2, 3.3, 3.6 and 3.7.

motor cortex layer and lower motor neurons layer. Each element represents a group of neurons with the same functionality. In a cortical layer, each element is analogous to a cortical column. There are six elements in the lower motor neurons layer, representing average activation of each of six muscles that controls upper and lower arm. The proprioceptive input layer consists of twelve elements, with six of them representing the length information of the six muscles, while the other six elements representing the tension information. The activation in the proprioceptive input layer is not governed by any activation rule, but by the muscle's geometric configuration (deciding muscles' length) and the activation of lower motor neurons (deciding muscles' tension). According to Equations 3.2, 3.3, 3.6 and 3.7, once the activation of a pair of muscles is decided, the corresponding length and tension information is uniquely determined.

The proprioceptive cortex layer (PI layer) contains 400 elements forming a 20 by 20 two-dimensional, hexagonally tessellated layer, with each element connected to its six neighboring elements. To avoid edge effects, elements on the edges are connected with elements on the opposite edges, forming a torus. These lateral connections are important in forming central excitatory, peristimulus inhibitory activation patterns. The proprioceptive input layer is fully connected to the PI layer. The motor cortex layer (MI layer) has the same size and structure as the PI layer. The PI layer is partially connected to the MI layer, with a coarse topographic ordering. That is, each element in PI is connected to its corresponding element in MI and the surrounding MI elements within a radius of four. This coarse topographic pattern of connectivity is motivated by previous experimental studies that have demonstrated topographic ordering



of excitatory connections from primary sensory cortex to MI [Asanuma, 1989; Jones *et al.*, 1978; Porter *et al.*, 1990; Yumiya & Ghez, 1984]. Also some previous studies in our research group indicated that the initial coarse topographic ordering could be refined during training under certain conditions [Reggia *et al.*, 1992; Sutton *et al.*, 1994]. From a computation point of view, such partial connections, as opposed to full connections between other layers, have greatly reduced the number of connections and make this model computationally tractable.

The lower motor neuron layer contains six elements representing the activation sent to the six muscle groups from MI. The MI layer is fully connected to the lower motor neuron layer. Weights on all of these interlayer connections are initially random. The transformation of muscle activation into proprioceptive information by the simulated arm effectively connects the lower motor neuron layer and proprioceptive input layer, and completes the closed loop system. In such a closed loop system, the activation of any layer will spread into subsequent layers and in this fashion influences itself. For instance, the activity of elements in the MI layer spreads to lower motor neurons, positions the arm, activates proprioceptive inputs, activates the PI layer, and thus ultimately changes the activation pattern in the MI layer.

In this network structure, the proprioceptive input layer and the lower motor neuron layer are greatly simplified, with each element representing a certain kind of information of an entire muscle. On the other hand, the cortical layers (both PI and MI) have many more representing elements and have a two dimensional structures. Since the purpose of this motor control model is to study the formation of cortical feature maps and their interrelationships, cortical network layers must be represented in enough detail for such purpose, while the input and output layers can be simplified to provide only the necessary biologically plausible information.

### 3.3 Activation and Learning Rules

The methods used in this network are competitive distribution of activation along with competitive learning. As illustrated in the previously, the competitive distribution of activation has some advantages in forming the central-excitation, peristimulus-inhibition responses (Mexican Hat response) that support map formation. Competitive learning is a widely-used unsupervised learning method.

The specific activation rule used is:

$$\frac{da_k(t)}{dt} = c_s a_k(t) + (max - a_k(t))(in_k(t) + ext_k(t)) \quad (3.8)$$

where

$$in_k(t) = \sum_j out_{kj}(t) = \sum_j c_p \frac{(a_k^p(t) + q)w_{kj}}{\sum_l (a_l^p(t) + q)w_{lj}} a_j(t). \quad (3.9)$$

Here  $c_s$  is the decay or self-inhibition constant indicating how fast the activation decays, and  $max$  is the ceiling value of activation. Constant  $c_p$  is the output gain constant, determining the fraction of activation to be output. Both parameters  $p$  and  $q$  can influence the degree of competition. The larger  $p$  is or the smaller  $q$  is, the more competitive the model. The value  $ext_k(t)$  is the external activation received by element  $k$ . This activation rule applies to the elements in the PI, MI, and the lower motor neuron layer. The activation in the proprioceptive input layer is not governed by this rule but by the arm mechanisms explained above.

The competitive learning rule used in this model is:

$$\Delta w_{kj} = \eta[a_j - w_{kj}]a_k^* \quad (3.10)$$

where

$$a_k^* = \begin{cases} a_k - \alpha & \text{if } a_k > \alpha \\ 0 & \text{otherwise} \end{cases} \quad (3.11)$$

and where  $\eta$  is a small learning constant. Only the weights from the arm layer to the PI layer are changed by Eq. 3.10; the cortico-cortical connections are constant. The value  $\alpha$  is the threshold and remains fixed throughout training. It ensures that only nodes with enough activation learn. This learning rule applies to all interlayer connections. Those intralayer connections in each cortical layers remain constant through out training process.

### 3.4 Experimental Methods

The experiments done with this model are divided into two parts. The first part involves the training process. Initially, all of the weights in the inter-layer connections were random, so the initial maps were poorly organized. Although the closed-loop has formed after the network was established, there was no activation in any network layer. The training starts by providing some external activation at some point of closed-loop network, so that this activation will circulate around the network structure to exhibit dynamics and flexibility of the model. The second part of each experiment involves measurement. After training is finished, the corresponding cortical feature maps are measured in certain ways in order to study the effect of training.

Training was done by stimulating the MI layer, i.e., by providing activation patches at randomly selected positions in MI. The system was driven by this initial stimulation and the subsequent activation was determined by the activation rule and feedback information via the closed loop system. Without clamping any element’s activation value, the system was able to get sufficient feedback information and no external influence (except the initial stimulation) was exerted on any layer in the system. The feedback information was able to influence the motor output, and eventually changed the feedback. Such a closed-loop of information (or activation) flow continued until the system achieved stablized activation levels in all of the layers, at which point the sensory feedback was fully consistent with motor output. Learning was conducted in an unsupervised fashion after such a stablized situation is achieved. The learning rule was applied to all weights of interlayer connections at the same time. This training process continued until well-organized cortical feature maps formed.

This training method is motivated by the presumed experiences of an infant exploring space without visual guide. An infant may initiate random activation patterns in motor cortex that result in arm movement. By associating feedback information received from the proprioceptive pathway and the motor commands issued, the cortex is able to self-organize. In this model, the initial stimulation to the MI layer represents input to MI from other, non-modeled brain areas. Since little is known about how other cortical areas issue motor commands to MI, a patch of activation was applied on different randomly selected positions in this model to simulate random movement commands received by MI.

More specifically, the training procedure was as follows:

- Step 1: Establish the network, forming a four layer, closed-loop system.

- Step 2: Randomly initialize connection weights for all inter-layer connections between 0.1 and 1.0.
- Step 3: Apply a patch of activation (radius 1, level 0.03) at a randomly selected position in the MI layer. This patch of activation is retained throughout the learning cycle as external input  $ext_k(t)$ , as indicated in Equation 3.8. The supplied input activation is combined with feedback activation from PI to jointly determine the activation in the MI layer.
- Step 4: Propagate the activation in MI to the lower motor neuron layer, using competitive distribution of activations, illustrated by Equations 3.8 and 3.9.
- Step 5: Compute the resultant joint angles, muscle length and muscle tension values of the model arm according to the transformation mechanisms described in the Equations 3.2, 3.3, 3.6 and 3.7; then use muscle length and tension values as activation values for elements in the proprioceptive input layer.
- Step 6: Propagate the proprioceptive input layer activation to and within the PI layer, using Equations 3.8 and 3.9.
- Step 7: Propagate the activation in PI to the MI layer and within the MI layer, using same activation rules.
- Step 8: Repeat Steps 4 through 7 for multiple iterations until the activation levels in each layer stabilize. Stabilization is determined to be a preset number of iterations (120) which was decided empirically by tracing the activation value for more iterations.
- Step 9: Use unsupervised learning to train the inter-layered connections (Eq. 3.10).
- Step 10: Repeat Steps 3 through 9, applying initial patch activation stimulation at different positions in MI, for a preset number of stimuli.

The convergence of training is determined in two ways. One way is to continue to train the networks for more learning cycles, to see whether further training causes qualitatively different maps. Usually, after a certain number of learning cycles, the maps that have formed in cortical layer exhibit certain characteristics and relationships. Further training could continue to generate gradually changed cortical feature maps, due to the randomness in the training, but all of the characteristics and their inter-relationships still remain unchanged. In this case, as is often done with models of this sort, the training process was considered completed. Another way to decide whether training has completed or not is to examine the input-output consistency of the system. This is done by stimulating (also clamping) the elements in the lower motor neuron layer one by one, and recording the corresponding activation pattern in MI. Then, these activation patterns are compared to the MI outgoing weights to the corresponding elements in the lower motor neuron layer. If they match well, then the system would be self-consistent. From a computational point of view, this kind of consistency indicates the convergence of training. A property of the unsupervised learning rule we used is that incoming weights will always shift to approximate the input activation patterns that activate a cortical element. Since the MI outgoing weights, which are the incoming weights of the lower motor neuron layer, matched well with the MI activations, the weights should no longer shift as long as the nature of input patterns does not change, and training has essentially

completed. If the system is not self-consistent, the weights would keep following the activation patterns, and qualitatively different cortical maps could still be generated.

After training is complete, the trained network is examined to see whether cortical feature maps have formed. The measuring of cortical feature maps are very similar to those in biological experiments. There are two kind of cortical feature maps for any cortical layer: input maps and output maps. For sensory input maps, we measured the cortical activities when certain sensory features are turned on. For example, to measure the input feature maps in PI with respect to proprioceptive inputs, one of the twelve elements in the proprioceptive input layer was tuned on, with the other eleven elements turned off, forming an activation pattern of the form  $(0, 0, \dots, P, \dots, 0)$ . This activation pattern was held steady (or clamped) while activation was propagated to the PI layer. The activation pattern in the PI layer was recorded after it stabilized. Such measurements were conducted for each of the twelve proprioceptive input features, all of whose corresponding activation patterns in PI were recorded for later analysis. For motor output maps, biological experiments usually involve stimulating certain locations in MI and measuring contractions of muscles either by movement perception or by EMG (electromyogram) [Donoghue *et al.*, 1992]. Similarly, the motor output maps in this computational model were measured by activating cortical elements, one at a time, and measuring the activation patterns in the lower motor neurons.

In this chapter, the motor control model was described in some details, as were experiment methods. The description has been focused on the common part of the model relevant to all of the work that follows. Since there are different variations of this model, it is inevitable that some different aspects of the model will be described in subsequent chapters. In Chapter 4, the motor control model with proprioceptive input alone will be described. This is the basic model, and its network structure is already described in this chapter. Chapter 4 will only add a more detailed descriptions that is not covered in this chapter and devote a major part to reporting simulation results. Chapter 5 will describe a variation of the motor control model with visual inputs only. Chapter 6 will describe the model with combined proprioceptive and visual inputs.

## Chapter 4

# Motor Control Model with Proprioceptive Input Only

The motor control model described in this chapter is the basic model, described in Chapter 3, that uses only proprioceptive inputs as sensory feedback to form a closed-loop system. This chapter will give specific parameter setting, simulation results, and the insights gained from these simulations.

### 4.1 The Model and Parameters

The network architecture, activation rule, and learning method in the version of the motor control model studied in this chapter have all been described in Chapter 3. The specific parameter values used in Equations 3.8-3.11 of the model in producing the results described in the next section are summarized in Tables 4.1 and 4.2. The learning threshold,  $\alpha$ , is 0 except  $\alpha = 0.32$  from MI to lower motor neuron layer. Selection of some parameters was motivated by our previous experiences with cortical modeling [Sutton *et al.*, 1994; Cho & Reggia, 1994]. Other parameters were obtained empirically in preliminary simulations so that three things were true: (1) the activation values of elements in each layer fell within reasonable ranges; (2) intracortical inhibition was sufficient for distinct features to emerge when maps were formed; and (3) a reasonable learning speed was achieved. For example, the relatively large value of  $q$  between the proprioceptive inputs and PI, and the large value of  $p$  between MI and lower motor neurons, allowed the input stimuli to MI to more quickly influence neurons immediately “downstream” in the closed-loop and more slowly influence more distant neurons in PI. This was found empirically to lead to much better map formation. Other parameters, such as  $c_s$ ,  $M$  and  $c_p$ , were set appropriately so that the activation level of elements was mostly between 0 and 1.

Although simulation results reported in next section are based on only one set of parameters, qualitatively similar results may be obtained from a variety of parameters values. In general, a small variation of any of the parameters will produce qualitatively similar results. For example, I found that using all zero learning thresholds gave similar results. More extreme variations of parameters may yield different maps, but maps generated by these variations may still preserve the general properties presented in this paper. For example, the lateral connectivity radius in cortical layers was increased from 1 to 2 or more. While this resulted in larger activation clusters in the cortical layer, the qualitative results presented in the next section still hold. Parameters used here are among those giving the best results we observed, but there is no guarantee that they are optimal. In general, the simulation results reported here are robust.

After training, the maps in different cortical layers were examined. These maps included the MI input and output maps, and the PI input and output maps. The measuring of maps was analogous

Parameters	PI layer	MI layer	Motor Neurons
$c_s$	-4.0	-2.0	-2.0
M	5.0	3.0	1.0

Table 4.1: Parameters used in activation update rule.

Parameters	Arm to PI	PI to PI	PI to MI	MI to MI	MI to Motor
q	0.1	0.0001	0.0001	0.0001	0.0001
p	1	1	1	1	2
$c_p$	0.8	0.8	0.6	0.4	0.05
$\eta$	0.2	NA	0.2	NA	0.1

Table 4.2: Parameters used in activation dispersal rule and learning rule.

to methods used in biological experiments. Generally, the input maps are measured by supplying different input stimuli and recording the cortical activations; the output maps are measured by stimulating cortical elements and recording the activations in the lower motor neuron layer. For each kind of map, there are two slightly different ways of showing it. One way is to represent an element with the kind of stimulus to which its response is the strongest. The second way is to show an element with all features that it responds to strongly (above a certain threshold). The first way emphasizes the most prominent feature, while the second way emphasizes multiple prominent features. The nature of a map is more clearly illustrated by using both kinds of map.

Length (tension)	Muscle
E (e)	upper arm <b>E</b> xtensor
F (f)	upper arm <b>F</b> lexor
B (b)	upper arm a <b>B</b> ductor
D (d)	upper arm a <b>D</b> ductor
O (o)	lower arm extensor or <b>O</b> penner
C (c)	lower arm flexor or <b>C</b> loser

Table 4.3: Labeling of muscle length and tension in illustrations.

## 4.2 Results

To illustrate the simulation results, the resultant input and output maps are illustrated, and then a comparison between input and output maps is given. The cortical elements that are tuned to multiple sensory features or control multiple muscles are also studied. The symbols used to represent map features are given in Table 4.3. For input maps, capital letters indicate cortical elements active when the corresponding muscle is stretched (increased length), while lower case letters indicate elements activated by increased tension in the corresponding muscle. For the output map, only capital letters are used to represent muscle contraction/activation.



Figure 4.1: Sensitivity of PI elements to muscle length before (left) and after (right) training (threshold=0.4). The numbers in the bottom represent the number of elements that are tuned to each of the six muscle length features in the same order as they are listed in Table 4.3.

#### 4.2.1 Input Map Formation and Characteristics

The measurement of input maps was done by applying twelve different activation patterns to the proprioceptive input layer, in each of which the activation of one of the 12 proprioceptive elements was non-zero, while for all the other elements it was zero. This is analogous to stimulating proprioceptive receptors from a single muscle group and measuring the resulting cortical activities. In this experiment, the input maps in both PI and MI layer were characterized.

Fig. 4.1 shows the PI input map, before (left) and after (right) training, using the symbols in Table 4.3. Each symbol in the map represents the feature to which the element in the corresponding location is most sensitive (i.e., the largest activation value above threshold). Those elements that responded below threshold to all inputs are represented as '-'. For example, the element in the upper left corner of the PI layer was not tuned above threshold to any specific muscle length or tension before training, but was tuned to upper arm adductor length (D) after training.

The map shown in Fig. 4.1 is difficult to understand. Fig. 4.2 shows the same PI input map in another way, illustrating only one type of elements that are tuned sufficiently strongly to a certain feature. Here Fig. 4.2 shows elements tuned sufficiently strongly to the length of the upper arm extensor (Fig. 4.2 (a) and (b)) and flexor (Fig. 4.2 (c) and (d)). Because this kind of figure gives a better indication of the distribution of the responding elements, it is used in the following, as long as there is no qualitative difference between different muscles. From Fig. 4.2, it is clear that after training, *elements tuned to the same proprioceptive feature formed clusters that are*

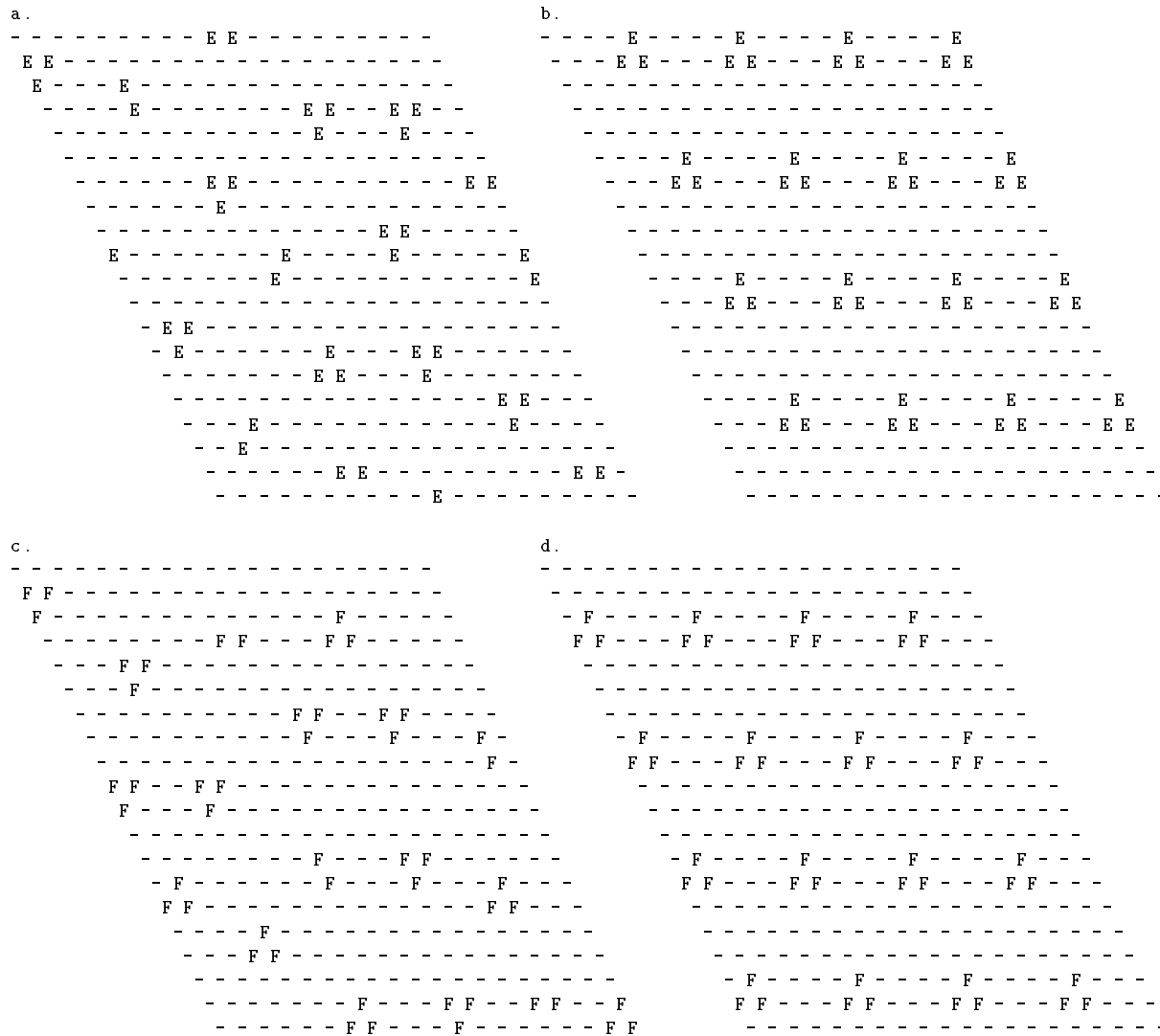


Figure 4.2: Tuning of PI elements to the length of the upper arm extensor (E) and flexor (F) before (left) and after (right) training (threshold=0.4).



*generally uniform in size and shape, and had centers arranged in a regular distribution.* The maps corresponding to other proprioceptive features showed similar qualities (See Appendix A.1.1 for a complete list of maps of all muscles). This kind of regularity indicates that a map has organized in the model PI layer with respect to proprioceptive features. The details of this map vary somewhat depending on the exact display threshold used (0.4 here), but the basic results remain the same. In addition, detailed study of the PI map in isolation show that although variation in intracortical lateral connection radius, intensity of lateral inhibition, and overall network size affect map details, the same qualitative results still hold [Cho *et al.*, 1994].

Figure 4.3 shows both the length and tension maps in the PI layer after training. By comparing these maps in the proprioceptive cortex layer, one can see that *the length map of a particular muscle matches well with the tension map of its antagonist muscle.* For example, the length map of the upper arm extensor matches the tension map of the upper arm flexor (Fig. 4.3 (a) and (d)), and the length map of the upper arm flexor matches the tension map of the upper arm extensor (Fig. 4.3 (b) and (c)). This type of relationship between length and tension maps is a result of training, i.e., it is not present prior to training. Since the activation of one muscle (increased tension) causes it to contract, thus stretching its antagonist muscle (increased length), there is a correlation between one muscle’s tension and its antagonist’s length in each input pattern. The maps capture the temporally correlated features of input patterns, reflecting the mechanical constraints imposed by the model arm.

The above paragraph described the relationship between different input feature maps. The alignment of feature maps is measured by visual comparison. There is another, more objective way to measure map alignment. Here we call it a similarity measuring method, as it quantitatively measures the similarity of two maps. With this method, the similarity of two feature maps is measured by taking the two corresponding activation patterns as vectors and calculating their normalized dot product (or inner product). For example, if features A and B (here A or B can be any of features indicated in Table 4.3) has corresponding activation patterns **A** and **B**, in vector format, then the similarity measurement of these two features is defined as:

$$\cos(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} \quad (4.1)$$

here the symbol ‘ $\cdot$ ’ is the ordinary dot product of vectors;  $\|\mathbf{A}\|$  and  $\|\mathbf{B}\|$  represent the length of vector **A** and **B**. In fact, Equation 4.1 simply calculates the cosine value of the angle formed by vector **A** and **B** in a 400 dimensional space. This value is always in the range of [0,1], because all the components of the vectors are non-negative. The similarity value becomes 1 when two vectors are in the same direction, in which case the maps of the two features are completely aligned. On the other hand, the similarity value becomes 0 when no component of both vectors has a non-zero value at the same time. In this case, the corresponding maps have no overlap at all. In the intermediate cases, the similarity values are somewhere between 0 and 1, and the corresponding feature maps are partially aligned.

Table 4.4 shows the similarity values between length and tension feature maps in proprioceptive cortex layer before training. In this table, each value is the similarity measurement of corresponding length and tension feature maps (using representing letters as illustrated in Table 4.3). Each column represents the correlations of different tension input features to the same length input feature, while each row represents the correlations of different length input features to the same tension input feature. For example, the value 0.20 in column F and row e is the similarity value between the

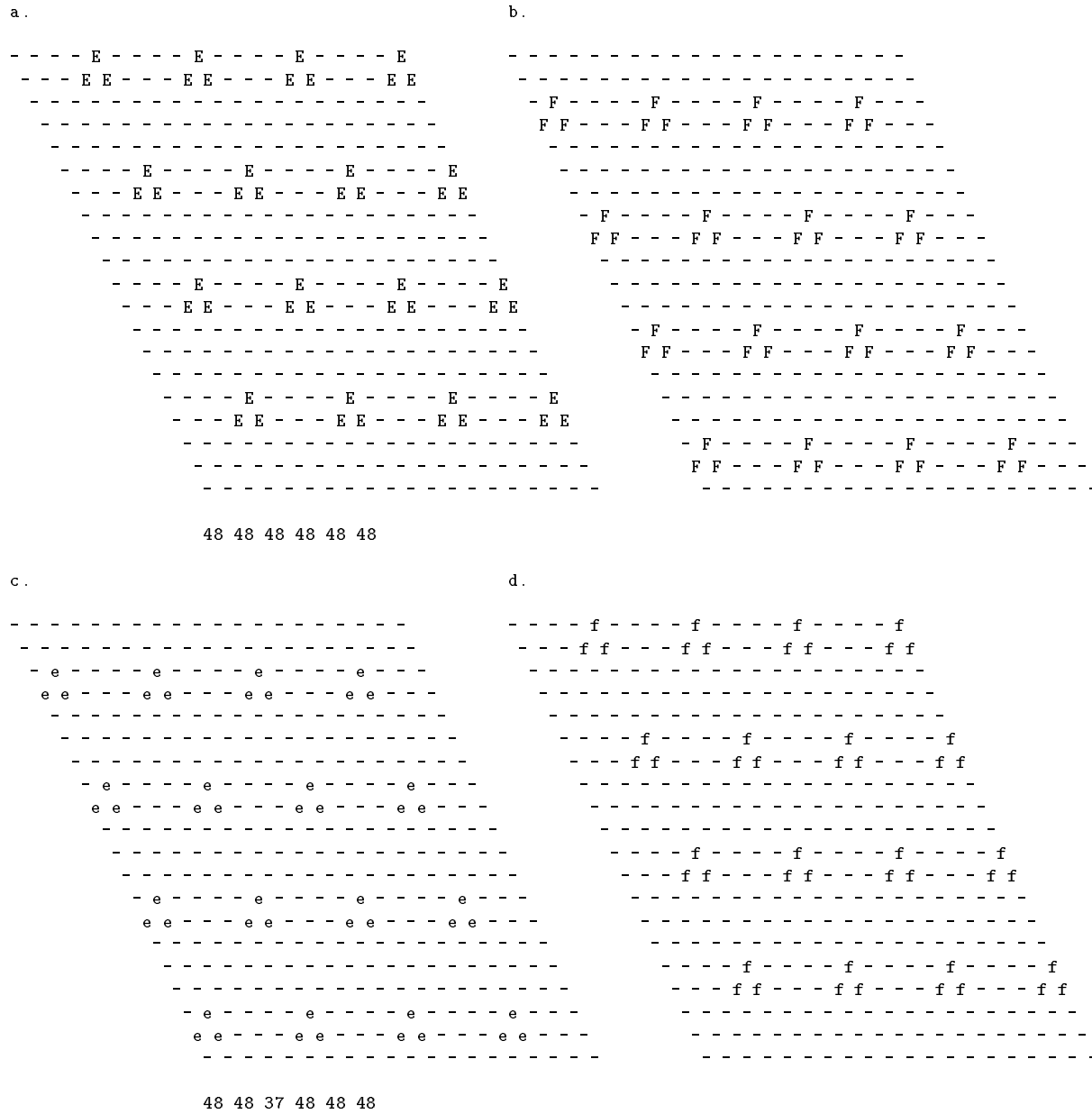


Figure 4.3: PI elements which are tuned above threshold to selected proprioceptive stimuli after training (threshold=0.4): (a) elements tuned to length of upper arm extensor, (b) elements tuned to length of upper arm flexor, (c) elements tuned to tension of upper arm extensor, (d) elements tuned to tension of upper arm flexor.

	E	F	B	D	O	C
e	0.18	0.20	0.10	0.17	0.31	0.11
f	0.10	0.15	0.25	0.28	0.18	0.42
b	0.14	0.15	0.12	0.17	0.26	0.19
d	0.07	0.20	0.17	0.13	0.25	0.13
o	0.16	0.22	0.07	0.27	0.19	0.18
c	0.12	0.14	0.22	0.19	0.22	0.32

Table 4.4: Similarity values between length and tension input features in proprioceptive cortex before training.

	E	F	B	D	O	C
e	0.00	<b>0.98</b>	0.06	0.00	0.18	0.32
f	<b>0.93</b>	0.00	0.12	0.14	0.00	0.01
b	0.03	0.00	0.00	<b>0.97</b>	0.04	0.01
d	0.08	0.01	<b>0.95</b>	0.02	0.21	0.06
o	0.07	0.30	0.01	0.01	0.00	<b>0.98</b>
c	0.03	0.04	0.04	0.02	<b>0.83</b>	0.00

Table 4.5: Similarity values between length and tension input features in proprioceptive cortex after training. Those values that are bigger than 0.7 are indicated in bold.

length feature of upper arm flexor (F) and the tension feature of upper arm extensor (e). In this table, none of the similarity values is bigger than 0.5, indicating that before training, there are no strong correlations between any pairs of input features.

Table 4.5 shows the same similarity values between length and tension feature maps in the proprioceptive cortex layer after training. In this table, those pairs that have strong similarity values (in bold face) are: (E, f), (F, e) (B, d), (D, b), (O, c) and (C, o) . It is apparent that the length feature of each muscle is strongly correlated with the tension feature of its antagonist muscle, and vice versa. This result is consistent with the visual comparison between input feature maps given above.

The similarity measuring method is an accurate, easily interpreted method. It gives a quantitative measurement of how similar two cortical feature maps are. This is more important when some cortical feature maps are only partially aligned, in which case it is difficult to draw conclusions from subjective visual comparisons. In the model described in this chapter, the cortical feature maps either tend to be aligned or do not overlap. So visual comparison is also effective (and most of the time, more intuitive) for studying the relationships between cortical feature maps. In the following chapters, some of the cortical feature maps are only partially aligned or overlapping. In those cases, the similarity measuring method show its advantages.

Figure 4.4 shows the length and tension maps in the MI layer after training The input maps in this layer undergo a transformation, when compared with the input maps in the PI layer. While clusters formed in this layer with a certain degree of regularity during training, it is apparent that the clusters in these post-training maps are less uniform in size, shape and periodicity, compared to the corresponding post-training PI input maps (Fig. 4.3). However, the same internal relationships still hold for the MI input map as for the PI input map: the length map of a particular muscle

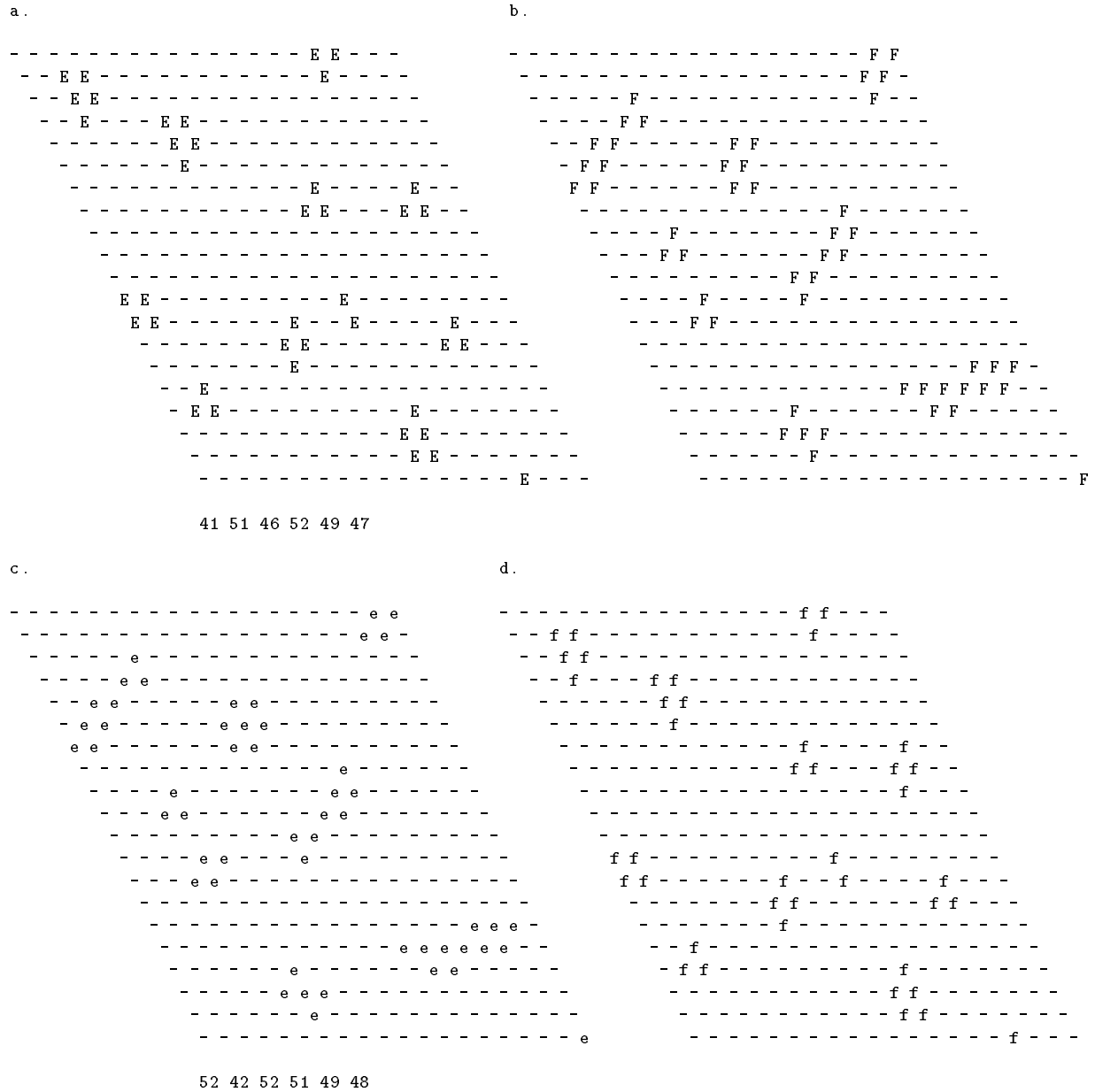


Figure 4.4: MI elements tuned above threshold to selected proprioceptive stimuli after training (threshold=0.4): (a) elements tuned to length of upper arm extensor, (b) elements tuned to length of upper arm flexor, (c) elements tuned to tension of upper arm extensor, (d) elements tuned to tension of upper arm flexor.

matches well with the tension map of its antagonist muscle. This indicates that the MI layer, although a step further away from the model arm where the mechanical constraints exist, still captures this feature of the input patterns.

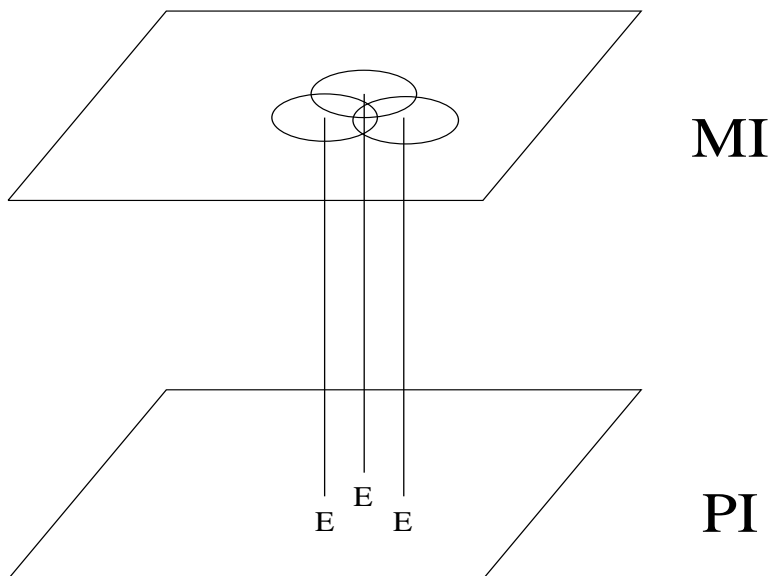


Figure 4.5: Intersection of activation areas. Each activated element in the PI layer spreads activation to its corresponding element in the MI layer and its nearby elements. The MI elements in the intersection area get most activation.

The fact that the PI input maps are quite different from the MI input maps has led to further observations about the transformation of maps from PI to MI. Although the connections from PI to MI initially have coarse topographic projections, such a topographic mapping was not refined during the training process, as was seen in the SI model described in Chapter 2 [Sutton *et al.*, 1994]. There are two reasons for this. First, multiple groups activating elements in PI project their activations to the corresponding MI area within a certain radius. Therefore the projecting areas from different groups tended to intersect. As a result, these intersecting areas received more activation and finally win the competition in attracting more activation (Fig. 4.5). Second, the elements in the MI layer serve for both receiving sensory input and sending motor output. Thus the sensory information MI receives is used to change the motor output commands, which in turn change the sensory feedback. In this process, it is necessary for the MI layer to find a compromise activation pattern that allows the output commands to be consistent with sensory feedback, in order to form stable activation patterns in all layers. Fig. 4.6 shows some of the incoming weight vectors of MI elements after training. Instead of forming a Gaussian shaped distribution as in the SI model, these weight vectors exhibit diversified distributions. This indicate that the transformation from PI to MI has become more complicated. There is no apparent correlation between the maps in PI and MI. Rather, the MI input maps are more associated with the MI output maps.

Once again, the similarity measuring method was applied here to quantitatively describe the relationships between MI input maps. Table 4.6 shows all of the similarity values between length and tension feature maps in motor cortex layer before (left) and after (right) training. Before training, the similarity values between length and tension feature maps ranges from 0.13 to 0.46,

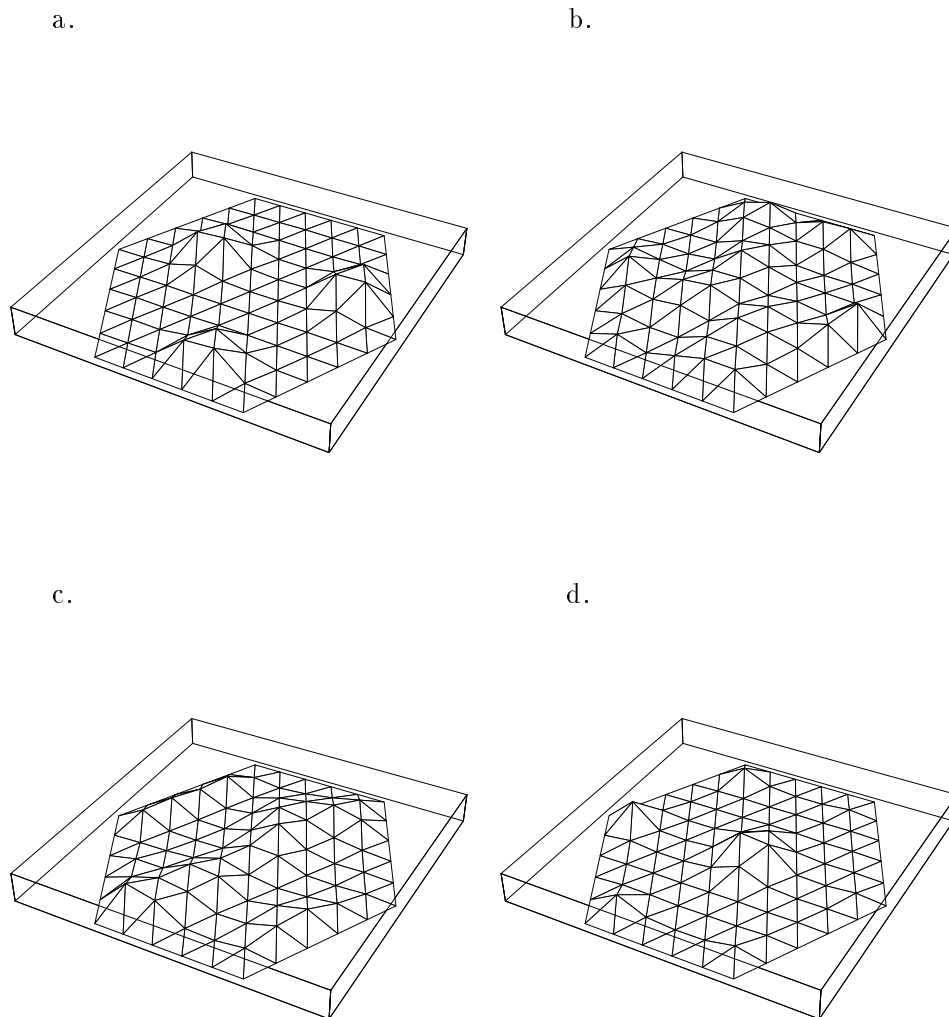


Figure 4.6: There is no single representative distribution of incoming weights in MI after training. Several different shapes are illustrated here. In each diagram, the width and length of the box represent the cortical surface, while the height represents the strength of the weight. The weight surface is plotted such that each weight is connected with its six neighboring weights. Therefore the distribution of weights is illustrated by the surfaces. (a) The incoming weights of node[2][2]: several peripherally located strong spots. (b) The incoming weights of node[0][10]: random-like shape. (c) The incoming weights of node[0][13]: stripe shape. (d) The incoming weights of node[1][6]: central strong shape with some scattered peripheral strong spots.

	E	F	B	D	O	C		E	F	B	D	O	C
e	0.30	0.24	0.16	0.13	0.40	0.21	e	0.00	<b>0.98</b>	0.02	0.02	0.07	0.12
f	0.30	0.34	0.35	0.33	0.31	0.46	f	<b>0.93</b>	0.00	0.03	0.03	0.01	0.01
b	0.24	0.19	0.15	0.35	0.32	0.23	b	0.00	0.01	0.00	<b>0.98</b>	0.01	0.01
d	0.20	0.34	0.20	0.26	0.36	0.25	d	0.03	0.01	<b>0.97</b>	0.00	0.13	0.01
o	0.27	0.27	0.20	0.38	0.18	0.25	o	0.03	0.11	0.02	0.01	0.00	<b>0.98</b>
c	0.29	0.20	0.24	0.24	0.20	0.40	c	0.01	0.01	0.00	0.02	<b>0.83</b>	0.00

Table 4.6: Similarity values between length and tension proprioceptive input features in the motor cortex layer before (left) and after (right) training. Those values that are bigger than 0.7 are indicated in bold.

indicating that no input feature map exhibits strong correlation with any other map. After training, the similarity values changed to be either very big (close to 1.0) or very small (close to 0.0). These values reflect the correlations between each length and tension feature, indicating that the length feature of each muscle is strongly correlated with the tension feature of its antagonist muscle.

#### 4.2.2 Motor Output Map Formation and Characteristics

The MI output map was examined after training by stimulating each MI element one by one and seeing which muscle(s) became activated. For simplicity, the MI output map was measured by examining the weights from MI to the lower motor neuron layer. In Chapter 7, some theoretical analysis will show that these two methods actually generate essentially the same feature maps, as long as appropriate threshold values are used. Fig. 4.7 shows the MI output weight map for the upper arm extensor muscle (E) and flexor muscle (F). Each ‘E’ (or ‘F’) means that the weight from MI to the lower motor neuron controlling the upper arm extensor (or flexor) is above a given threshold. Maps for other muscles show similar features (See Appendix A.1.3 for a complete list of maps of all muscles).

Comparing Fig. 4.7 (a) and (b), (c) and (d), it is apparent that clusters formed during training. These clusters are larger and more irregular than those in the PI input map. Some clusters suggest a tendency to form stripes. Although these clusters are not uniform in size and shape, they are similar to the activation patterns actually seen in mammalian MI cortex [Donoghue *et al.*, 1992].

#### 4.2.3 Consistency Between Input and Output Maps

In the previous paragraphs it was shown that the appearance of the MI input map is quite different from the PI input map (compare Fig. 4.3 and Fig. 4.4), and the projection from the PI layer to the MI layer is not in the previously expected topographic order. This raises the question of the nature of the relationship of the MI input map to the MI output map.

Fig. 4.8 (a), (b) shows the MI output maps of the upper arm extensor and flexor, marked by E and F, respectively, based on the MI output weights. Fig. 4.8 (c), (d) shows the MI proprioceptive maps with regard to the length and tension of the upper arm extensor, respectively, based on the activation of MI elements (above threshold) when the length or tension feature is present in the proprioceptive input layer. By comparing the MI output and input maps, it is seen that *the MI input map of a particular muscle’s length matches well with the MI output map of its antagonist muscle, while the MI input map of a particular muscle’s tension matches well with the MI output*

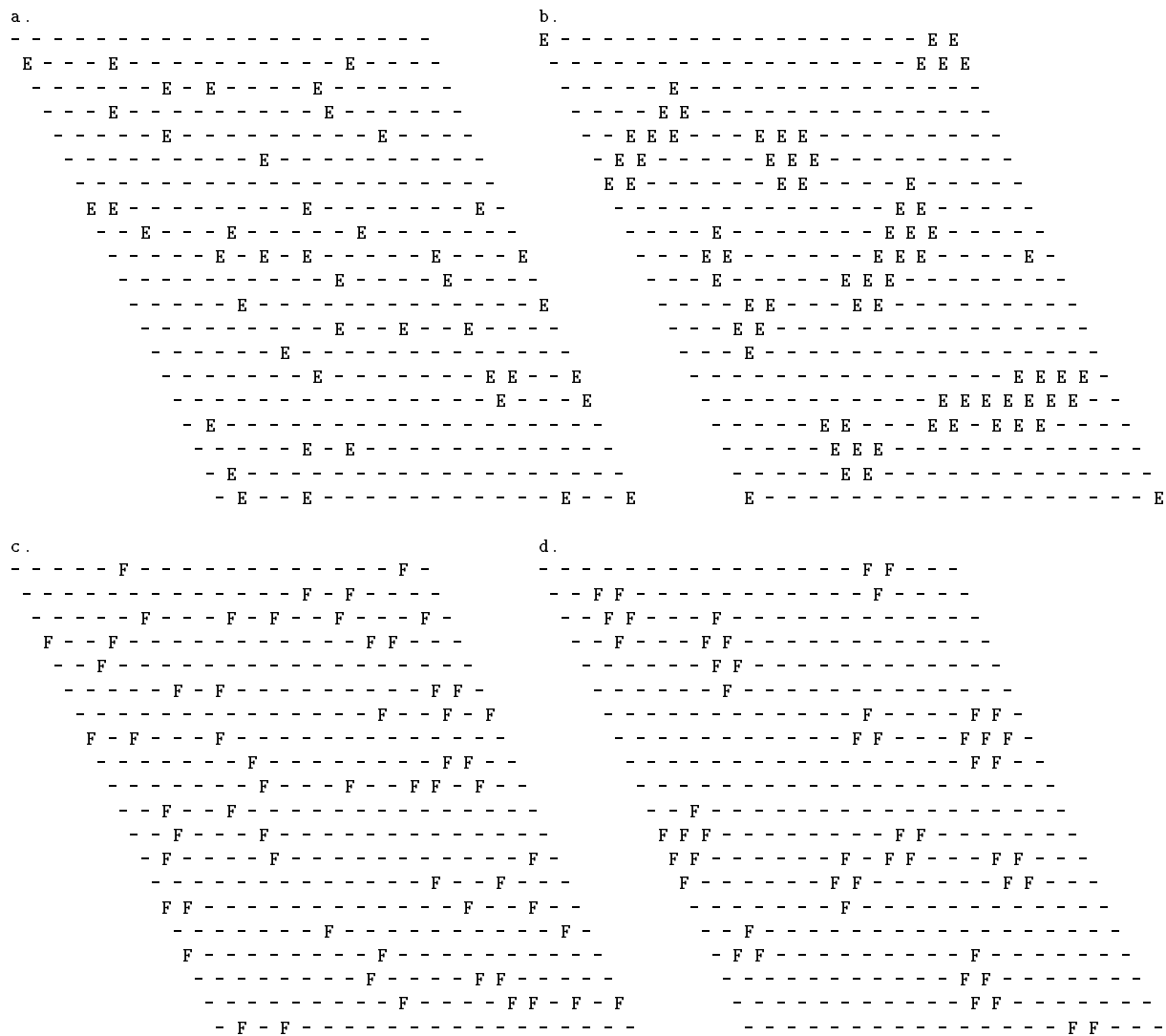


Figure 4.7: MI output map before (left) and after (right) training for upper arm extensor (E) and flexor (F) (threshold=0.4).



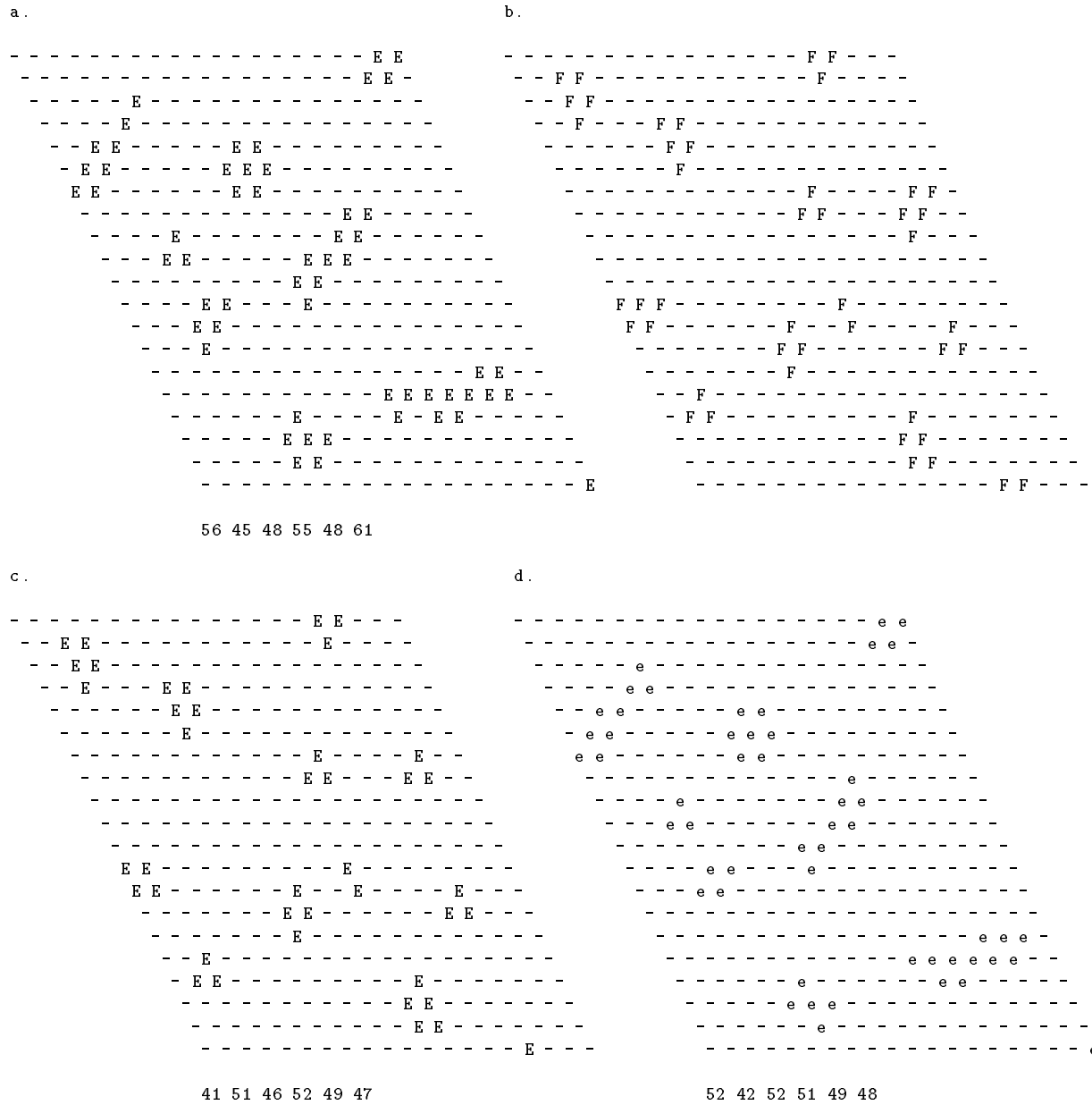


Figure 4.8: Comparing post-training MI output maps (threshold = 0.7) of the upper arm extensor (a) and flexor (b) with MI input maps (threshold = 0.4) of length (c) and tension (d) of the upper arm extensor.

	E	F	B	D	O	C		E	F	B	D	O	C
E	0.52	0.48	0.49	0.52	0.51	0.53	E	0.02	<b>0.95</b>	0.07	0.09	0.08	0.07
F	0.50	0.49	0.50	0.48	0.51	0.52	F	<b>0.96</b>	0.02	0.03	0.04	0.17	0.09
B	0.49	0.47	0.47	0.46	0.49	0.49	B	0.05	0.09	0.04	<b>0.98</b>	0.04	0.05
D	0.47	0.49	0.47	0.49	0.50	0.48	D	0.04	0.10	<b>0.97</b>	0.04	0.04	0.08
O	0.49	0.50	0.47	0.48	0.47	0.48	O	0.25	0.05	0.06	0.23	0.02	<b>0.86</b>
C	0.48	0.48	0.50	0.49	0.48	0.48	C	0.18	0.04	0.04	0.04	<b>0.98</b>	0.03
e	0.50	0.49	0.49	0.53	0.50	0.47	e	<b>0.95</b>	0.02	0.02	0.03	0.10	0.08
f	0.49	0.51	0.52	0.53	0.53	0.49	f	0.02	<b>0.97</b>	0.09	0.05	0.03	0.05
b	0.48	0.50	0.49	0.49	0.45	0.46	b	0.02	0.06	<b>0.95</b>	0.04	0.03	0.08
d	0.49	0.48	0.48	0.45	0.50	0.48	d	0.03	0.05	0.04	<b>0.97</b>	0.03	0.05
o	0.47	0.45	0.50	0.50	0.50	0.48	o	0.10	0.03	0.03	0.03	<b>0.98</b>	0.03
c	0.48	0.48	0.51	0.49	0.47	0.48	c	0.12	0.05	0.08	0.06	0.03	<b>0.96</b>

Table 4.7: Similarity values between length and tension input features in motor cortex before (left) and after (right) training. Those values that are bigger than 0.7 are in bold style.

*map of its corresponding muscle.* For example, the MI proprioceptive length map of the upper arm extensor matches well with the MI output map of the upper arm flexor (compare Fig. 4.8 (c) with (b)); the MI proprioceptive tension map of the upper arm extensor matches well with the MI output map of the same muscle (compare Fig. 4.8 (d) with (a)). The reason for this is that when a muscle is activated in producing a movement, it contracts, and its length typically decreases, while its antagonist muscle’s length is increased accordingly. At the same time, the activated muscle is under increased tension. Therefore activation of a muscle typically generates proprioceptive feedback indicating increased stretch of its antagonist muscles, and increased tension of itself. This kind of correlated activation of muscle length and tension feedback is captured by the model and reflected in the maps, such as those in Fig. 4.8.

Table 4.7 summarized the similarity values between MI input and output features before (left) and after (right) training. In each column of the table, the values correspond to the same output feature, represented by a capital letter on top of the column. In each row of the table, the values correspond to the same length or tension input feature, represented by a capital or lower case letter (as illustrated in Table 4.3), respectively, on the left hand side of the row. For example, in the table on the left hand side, the value 0.49 in the row of ‘E’ and column of ‘B’ is the similarity value between the length input feature of upper arm extensor (E) and the motor output feature of upper arm abductor (B) before training. In the same table, the value 0.50 in the row of ‘e’ and the column of ‘E’ is the similarity value between the tension input feature of upper arm extensor (e) and the motor output feature of upper arm extensor (E). It should be noted that the labels on top of the columns represent motor output features, and should not be confused with the label on the left hand side of each row, which represent input features. From Table 4.7, it is clear that before training the similarity values ranged randomly from 0.45 to 0.53, showing no strong correlations. After training, the similarity values changed dramatically to reflect the correlations of the self-organized feature maps. The upper half of the table on the right hand side illustrates the correlations between the length input feature and the motor output feature after training. We can see that the pairs of antagonist muscles have strong correlations, with similarity values from 0.86 to 0.98. The lower half of the table on the right hand side illustrates the correlations between the tension input feature and the motor output feature after training. It is quite clear that all of

the large similarity values (from 0.95 to 0.98) are on the diagonal line, indicating that the tension input feature and the motor output feature of the same muscle are strongly correlated. All these properties are natural results of the training.

#### 4.2.4 Elements Tuned to Multiple Features

In both the PI and MI layers, there are elements which became tuned to multiple proprioceptive input features. Some of these tunings are potentially incompatible with the constraints imposed by the mechanics of the model arm. For instance, it seems unlikely that a PI element would be tuned to both a muscle’s length and to its tension together since a muscle does not usually contract (high tension) and lengthen simultaneously in the model arm. Another implausible case would be that a PI element is tuned to the stretch of two antagonist muscles, since they cannot be stretched at the same time. Table 4.8 shows the number of PI and MI elements tuned to implausible pairs of inputs before and after training. On the first line of the table, implausible tuning pairs are given using the symbols in Table 4.3. For example, label (E,F) indicates that a cortical element is tuned to the length of the upper arm extensor and flexor simultaneously, i.e., that it is activated above threshold when either of these muscles is stretched. Following each label in the same column are the numbers of cortical elements that are tuned to the indicated pair of features. The number of PI and MI elements tuned to implausible pairs decreased to zero during training. This is clear evidence that the model learned the correlations between proprioceptive features arising due to the constraints of the model arm. It should be noted that the plausibility of map features here is predicated on the specific details of the model arm used. Thus, maps in our model would be unable to capture some correlations between muscle tension and length occurring with real movements. The key point here is that the feature maps do not represent implausible relationships for the given arm model.

Tuning Pairs	(E,F)	(B,D)	(O,C)	(E,e)	(F,f)	(B,b)	(D,d)	(O,o)	(C,c)	sum
PI before training	5	7	2	8	6	3	7	3	6	47
PI after training	0	0	0	0	0	0	0	0	0	0
MI before training	9	4	11	12	7	9	14	9	11	86
MI after training	0	0	0	0	0	0	0	0	0	0

Table 4.8: Numbers of implausibly tuned PI and MI layer elements (threshold=0.5).

MI elements that control multiple muscles were also examined. Fig. 4.9 shows the MI elements which have strong connections to multiple muscles after training. At a threshold of 0.4, there were, among 400 elements, 90 elements having strong connections to multiple muscles, and 16 of them controlled 3 muscles. With a higher threshold, the number of multiple control elements decreases. A careful examination of these multiple controlling elements shows that *most of them (85 out of 90) control muscles acting along different coordinates*. For example, as shown in Figure 4.9, the element in the second row and sixth column can activate the upper arm extensor (E), upper arm abductor (B) and lower arm extensor (O), each being one of the three pairs of antagonist muscles. This type of element is capable of producing coordinated movement of the arm toward a particular direction, in this specific case toward the upper back part of space. This result is consistent with the observation that some neurons in motor cortex code for movement direction. It also provide a testable prediction on the controlling of muscles by motor neurons that could be verified by biological experiments. It should be pointed out that in the mammalian motor system, the control of multiple muscles by individual MI neurons can be implemented via lower brain and

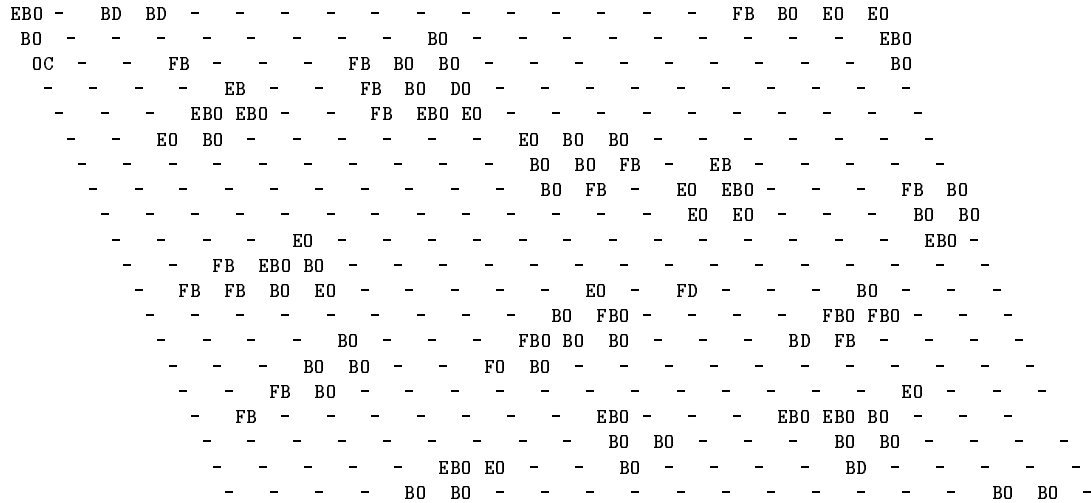


Figure 4.9: Map of MI elements strongly activating multiple muscles (threshold=0.4)

spinal circuitry, so our model is by no means analogous to biological systems in terms of actual neural circuitry. This result indicates that, via training, it is possible to produce this type of multiple muscle control in a more general sense from initially random connections.

#### 4.2.5 Sensitivity to Simultaneous Weight Adaptation

Our model makes the assumption that all three sets of connection weights (sensory neurons to PI, PI to MI, and MI to lower motor neurons) mature simultaneously. While relatively little is known about the precise development of these connections, there is some evidence that the PI to MI connections develop later than the others [Bruce & Tatton, 1980], and thus the developmental assumption in our model should be viewed as only a first approximation to reality.

To examine this issue, we undertook a single simulation with the same parameter values used in the simulation described above, where training was done sequentially. Specifically, we allowed sensory connections to learn first (2000 iterations), then those connections plus MI to lower motor neuron connections to learn (2000 iterations), then all connections to learn (2000 iterations), motivated by data in [Bruce & Tatton, 1980]. We used a smaller learning rate  $\eta = 0.05$  on sensory connections to PI to compensate for its longer total training time. The maps obtained and their alignments were qualitatively similar to those described above, although for one of the six muscle length inputs the alignment with motor outputs was not precise. We believe that a substantial joint learning phase is necessary for complete map alignment to occur. This result, plus the fact that qualitatively similar maps appear in isolated PI when it is trained by randomly positioning the arm [Cho & Reggia, 1994], suggests that the results obtained here are not sensitive to the exact developmental order of connection maturation.

### 4.2.6 Lesioning Study of the Motor Control Model

After development, the motor control model reported in this chapter was used to study the effect of lesions to the cortical feature maps [Goodall *et al.*, 1997]. After training was completed and the maps formed, the motor control model was subjected to simulated sudden, focal lesions to a cortex region. There are two sets of lesions studied: lesions of PI and MI. In both cases, an area of cortical elements were clamped to zero to simulate a lesion. The cortical feature maps were examined immediately after lesioning, as well as after the model was further trained with additional 2000 training patterns. These maps were compared with the corresponding maps of the pre-lesion model and a control model, which was being trained without being lesioned.

It was found that immediately after a PI area was lesioned, the perilesion area in PI became less active in responding to proprioceptive stimuli, forming a functional impairment zone. After further training, this functionally impaired zone became larger. On the other hand, lesioning of an area in MI caused perilesion regions to have increased activation. Further training increased the activity in these perilesion regions. These simulation results suggested that there are two phases of cortical map reorganization: a rapid reorganization in response to the focal lesion, and a slower reorganization after further learning. Also, the activity in the perilesion area was found to play an important role in long term map reorganization. Appendix B gives a more detailed description of the simulation of lesion study. It illustrates one way in which the motor control model can be used to study hypotheses about neurological disorders.

## 4.3 Discussion

Self-organizing feature maps have become important neural modeling methods over the last several years. They have not only shown great potential in application fields such as motor control, pattern recognition, optimization, etc [Ritter *et al.*, 1989; Kohonen, 1989; Angeniol *et al.*, 1988], but have also provided insight into how the mammalian brain becomes organized [von der Malsburg, 1973; Linsker, 1986; Grajski & Merzenich, 1990; Burnod *et al.*, 1992; Weinrich *et al.*, 1994; Sutton *et al.*, 1994]. The computational motor control model described here falls into this second category. It exhibits properties that are consistent with experimental findings involving biological motor control systems. It also provides us with knowledge about the organizing and processing of sensory and motor information along the input-output pathway. Some properties of the model are summarized as follows.

First, this model has shown spontaneous emergence of multiple feature maps during unsupervised learning. The model self-organized from initially random connections. These results indicate that, although this model is a significant simplification from reality, it has captured the basic structure and some principles of biological motor control systems. The fact that the model self-organizes into multiple feature maps that are stable in spite of its closed-loop nature suggests that the underlying assumptions (network connectivity, activation dynamics, unsupervised learning, etc.) can account for some important aspects of proprioceptive and motor map formation in mammalian cortex. We believe that these map formation results do not depend significantly on the specific form of the activation rule used in the model (Eq. 3.8 and 3.9), as long as a clear cut Mexican Hat pattern of lateral interactions occurs in the cortex [Reggia *et al.*, 1992]. For example, qualitatively similar results have been obtained when previous cortical map formation experiments using activation rules similar to those used here [Cho & Reggia, 1994; Sutton *et al.*, 1994] were re-implemented

using more standard activation functions.

Second, the maps formed capture the mechanical constraints of the simulated arm. Analysis of the proprioceptive input maps showed that the same elements were tuned to the length/stretch of a particular muscle and to the tension of its antagonist muscle. This is true for both cortical layers. It indicates that cortical elements are capable of recognizing the temporal correlations in the input patterns. This model also showed a consistent relationship between the proprioceptive input maps and motor output maps. It was found that the set of MI elements that control a particular muscle usually respond to the tension of this muscle and the length of the antagonist muscle. These results are biologically plausible and support the following hypothesis: *when multiple feature maps exist in the same region of cortex, features in one map that are temporally correlated with those in another will come to occupy the same spatial locations.*

Third, the motor output map generated in this model qualitatively resembles the map in mammalian motor cortex. Many experiments have been conducted on mammalian motor cortex, one of which is a systematic mapping of primate forelimb motor cortex [Donoghue *et al.*, 1992]. In that experiment, several major findings indicated that the organization of motor cortex is more complicated than previously thought:

- Property 1. Neurons representing the same muscle form separated, widely distributed clusters.
- Property 2. The size and shape of clusters representing the same muscle differ significantly from muscle to muscle, from subject to subject.
- Property 3. Many neurons in motor cortex can activate multiple muscles.
- Property 4. No apparent topographic relationships were found in the forelimb area of motor cortex.

Properties 1 and 4 are apparent in our computational model. Property 2 is also apparent when comparing the regularity of the proprioceptive input maps in PI with the irregular motor output maps in MI (compare Fig. 4.3 and Fig. 4.8 (a), (b)). Property 3 emerges in our model via unsupervised learning. As indicated by Fig. 4.9, many MI elements control multiple muscles. Also, by increasing the measurement threshold the number of MI elements that control multiple muscles decreased. This is also consistent with experimental data showing that stronger stimulation tends to recover more multiple-muscle neurons [Donoghue *et al.*, 1992].

This computational motor control model also provides testable predictions that can be verified or refuted by future biological experiments, as follows. First, to my knowledge, there has been no systematic mapping conducted on mammalian proprioceptive cortex. Thus, the characteristics shown in the cortical proprioceptive input maps, such as regular clusters of elements tuned to the same muscle tension/length, represent testable predictions, although we would not expect as precise regularity as occurs in our simplified model. The relationship between muscle length and tension features are also yet to be verified in biological experiments.

This motor control model also shows that proprioceptive sensory maps formed in the MI layer after training. These latter maps exhibit the same properties as seen in the PI layer. On the other hand, the proprioceptive maps in the MI layer, although they capture the same constraints, differ from the maps in the PI layer in terms of cluster size and shape. Analysis of the model reveals that this is due to the weights on connections from the PI layer to the MI layer. Even though the connections from the PI to the MI layer were initially coarsely topographic, training did not

refine this topographic projection, and the resultant weights became complicated and could not be characterized by any simple property.

The view that neurons in MI code for the force of exertion of individual muscles is controversial. Some of the neurons in MI activate multiple muscles [Donoghue *et al.*, 1992], suggesting that these neurons might code for movement direction rather than individual muscles [Georgopoulos *et al.*, 1986]. With this computational model, it was found that most multiply tuned MI elements controlled muscles in different muscle-group pairs and thus their activation tends to move the hand toward a particular direction. This finding is consistent with biological experiments showing that motor neurons tend to project to synergistic muscles [Cheney & Fetz, 1985] and with demonstrations that neurons in motor cortex code for movement directions [Georgopoulos *et al.*, 1986]. Experimentally, stimulation of motor cortex neurons tends to excite one muscle and inhibit its antagonist muscle, thus causing synergistic movements [Cheney & Fetz, 1985]. Whether the activation of motor cortex neurons activates muscles in different joints (or the same joint but different movement dimensions) is an interesting issue for future biological experiments. It should be noted that our computational model is built without *a priori* discrimination between muscles with respect to being antagonists or operating at differing joints; it is the training that distinguishes the muscles in different pairs.

## Chapter 5

# Motor Control Model with Only Visual Input

In this chapter, a variation of the motor control model is studied to investigate the effects of visual input as feedback. In biological motor control systems, the motor cortex receives input from afferent pathways other than proprioceptive input, including visual input [Johnson, 1992]. Most previous models involving visual input and motor output have tried to minimize errors in reaching movements [Kuperstein, 1988; Ritter *et al.*, 1989; Walter & Schulten, 1993]. In other words, these past models used learning rules in order to minimize the difference between target position and hand position in their sensory feedback. Little effort has been taken to analyze visuo-motor cortical feature maps that self-organize using unsupervised learning. In this section such an analysis will be conducted. The model still has a simulated arm, and a two dimensional layer of MI elements. Instead of having a proprioceptive input layer, however, a layer of visual inputs supplies feedback information.

### 5.1 The Model

Fig. 5.1 shows a schematic diagram of the model. In this model, the MI layer is again a 20 by 20 two-dimensional, hexagonally tessellated layer, with each element connected to its six neighbors. This layer is fully connected to the lower motor neuron layer, which has six elements representing activations of six muscles. The visual input layer has nine elements, which are fully connected with the MI layer. Visual input here represents an abstraction of hand position in shoulder-centered coordinates directed to MI from visual cortical regions [Johnson, 1992]. The transformation from lower motor neuron activation to hand position is based on the mechanism of the model arm described earlier, so this version of the model again forms a feedback loop.

In order to measure the visual input in space (i.e. hand position), it is necessary to define a coordinate system. Fig. 5.2 shows a coordinate system in relation to a human body which is facing into the page. The origin of the coordinate system is the shoulder of the right arm (where the simulated arm is anchored), and the positive directions of the coordinate axes are shown.

The nine elements in the visual input layer are divided into three groups. Each group has three elements, coding the hand position in one of the three dimensions: X, Y, or Z. In each dimension, the movement range of the hand position of the model was linearly scaled into  $[-1, +1]$ . The three elements coding the same dimension overlap but are tuned maximally to the negative, middle, or positive part of the range, respectively. The actual tuning formula (taking the X dimension as an



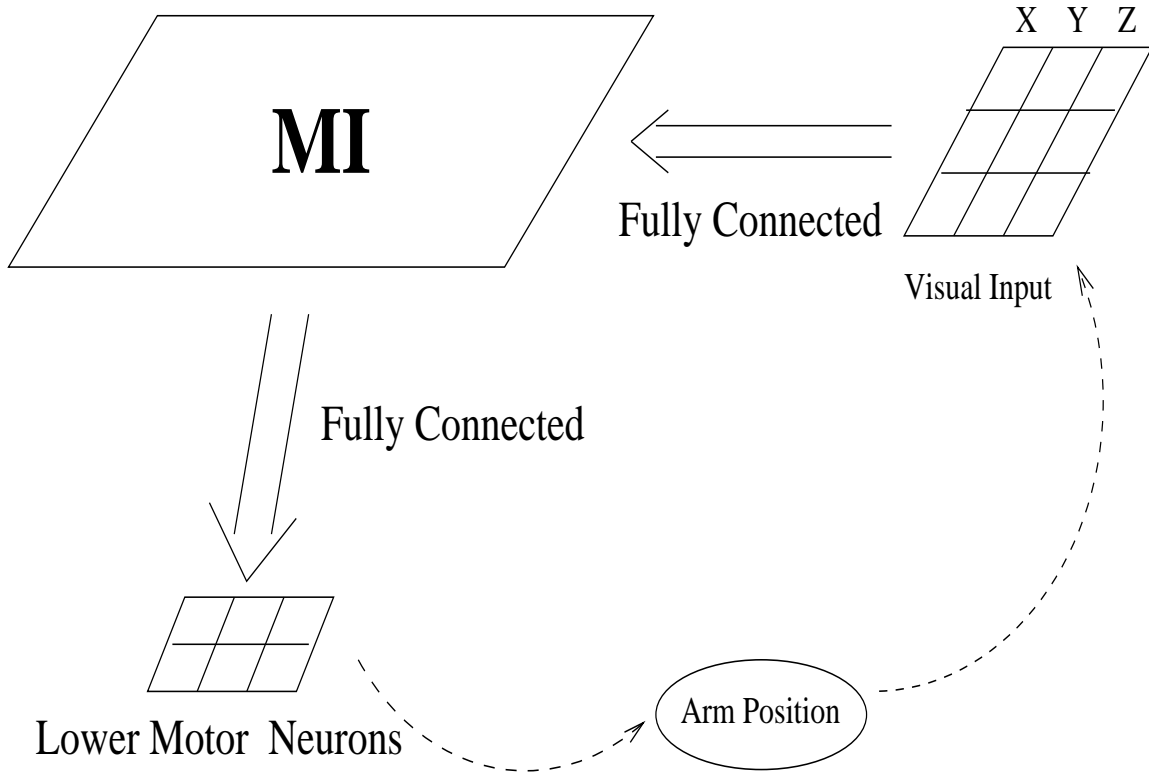


Figure 5.1: The motor system with visual inputs. The MI elements send activation to lower motor neurons, which direct the arm movement. There is a visual input layer, which receives coded information about the hand position of the model arm and supplies this information to the MI layer. The feedback to the MI layer influences the MI output and thus forms a closed-loop.

example) are:

$$act_{X1} = \max(-H_x, 0) \quad (5.1)$$

$$act_{X2} = \max(1 - 2 * \text{abs}(H_x), 0) \quad (5.2)$$

$$act_{X3} = \max(H_x, 0) \quad (5.3)$$

where  $H_x$  is the hand position in the X dimension that ranges from -1 to +1. Variables  $act_{X1}$ ,  $act_{X2}$  and  $act_{X3}$  are activations of the elements tuned to the negative, middle and positive ranges of the X dimension, respectively. Function  $\max$  takes the maximum of its parameters, while  $\text{abs}$  stands for absolute value returning the magnitude of its parameter. Fig. 5.3 shows graphically the tuning curves based on the above formulae. This coding scheme ensures that no matter what X value the hand position has (assuming it is in the range of [-1,+1], after scaling), there is always at least one visual element(s) activated. The three element coding is relatively coarse, and the overlap of the tuning ranges ensures a unique tuning pattern for every position. A similar representation of hand position is used for the Y and Z dimension.

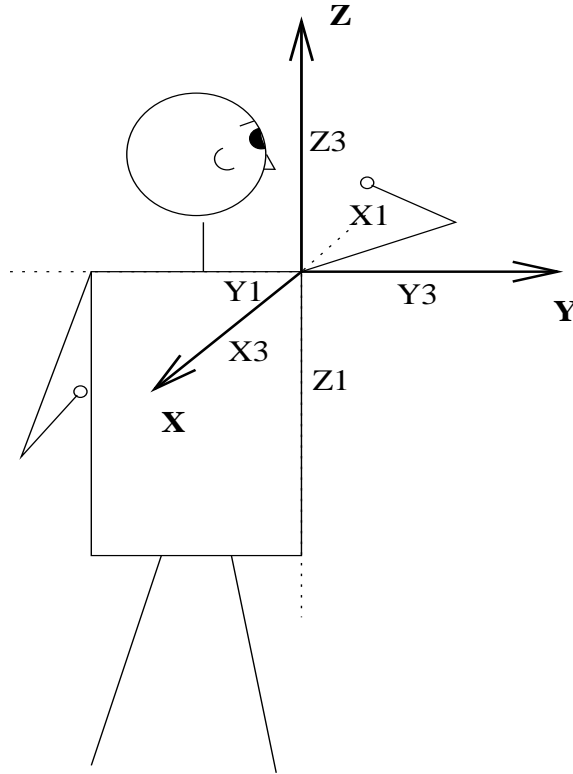


Figure 5.2: The coordinate system in the arm movement space relative to a human body that is facing into the page. The origin of the coordinate system is the right shoulder of the body, which is also one end of the simulated arm. The positive direction of the X axis is to the back of the body (out of the page). The positive direction of the Y axis is to the right side of the body. The positive direction of the Z axis is up. Small circles designate “hand” positions.

## 5.2 Experimental Methods

The training method is similar to that previously described. All the weights (except those intralayer connections in the MI layer) were randomly initialized. Training was done by applying a patch of activation (of radius 1) at randomly selected MI regions. The activation spread to the lower motor neuron layer, using the mechanism of competitive distribution of activation (Equation 3.8 and 3.9). The activation pattern in the lower motor neuron layer was then transformed into the corresponding activation pattern in the visual input layer, using the arm mechanism and the Equations 5.1-5.3. The visual input layer then spread the activation back to MI, and influenced the MI activation pattern. This formed a closed-loop system. No element’s activation was clamped in any layer. After sufficient time steps (120 were used empirically in this experiment), stabilized activation levels were achieved in all the layers. Learning was then conducted by applying the competitive learning rule to all of the inter-layer connections (Equation 3.10 and 3.11). The model was trained for 6000 learning cycles before the maps were examined. Further training would change the appearance of the maps but all the characteristics of the maps reported in later sections remain.

The parameter values used in Equations 3.8-3.11 of the model in this experiment are summarized in Tables 5.1 and 5.2. The learning threshold,  $\alpha$ , is 0 in all layers. The parameters used in this

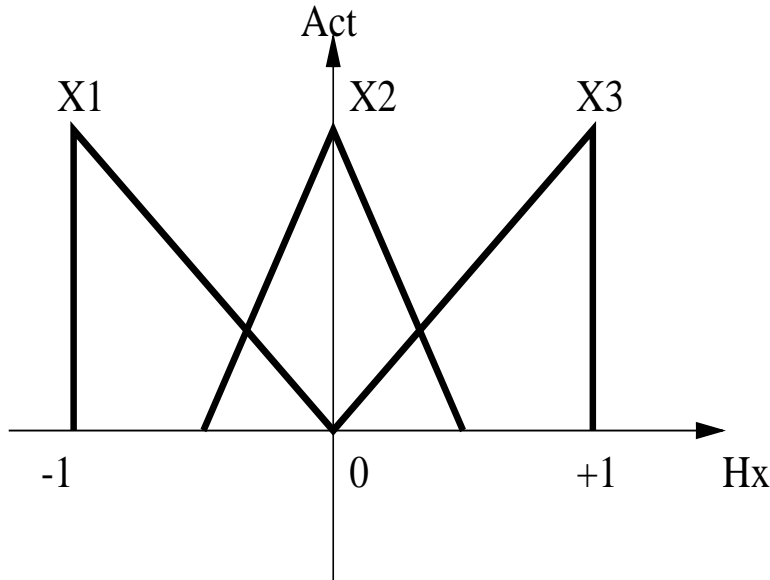


Figure 5.3: The tuning curves of the three visual elements coding hand positions in the X-dimension.

experiment are similar to those in previous experiments. Any small change to any parameter would not change the qualitative characteristics of the results discussed in the following sections.

After training, the MI input and output maps, along with their relationships, were examined. The MI output maps were measured in the same way as described earlier. Basically, for each of the six muscles, there is a corresponding MI output map that shows the MI elements with connections strong enough (above a certain threshold) to this muscle. The MI input maps with respect to the visual input were measured by stimulating visual input elements and measuring the corresponding MI activations. This was achieved by activating one of the nine elements in the visual input layer each time and holding that pattern steady. The corresponding MI activations were then examined after activation stabilization was achieved.

In the analysis of this model, in addition to examining the characteristics of the individual maps, it is also interesting to study the relationships of the input maps and output maps. Basically, we again are studying the correlation of different features, because each input or output map represents the distribution of a particular feature. There are two ways to study the correlation of features: by visual comparison and by using the similarity measuring method. Both methods have been used in the motor control model with proprioceptive inputs, as described in previous chapter, but need further consideration here because of the more complex relations between MI visual input and MI motor output maps.

The visual comparison method is more intuitive, especially when the two features investigated have a strong correlation or no correlation at all. In these cases, it is easier to see the relationship of the two features by visual inspection. However, this method has some limitations. First, it is a qualitative and subjective measurement. This method can only give a rough estimation about whether or not the two investigated features are correlated. It will not indicate how strong the correlation is. Second, some of the features have no clear one-to-one correspondence. That is, one feature could be partially correlated with multiple features. In this case, one map is partially aligned with multiple maps. It is difficult to use a visual comparison method to examine such

Parameters	MI layer	Motor Neurons
$c_s$	-2.0	-0.2
M	3.0	1.2

Table 5.1: Parameters used in activation update rule.

Parameters	Visual to MI	MI to MI	MI to Motor
q	0.0001	0.0001	0.0001
p	1	1	2
$c_p$	5.0	0.4	0.005
$\eta$	0.1	NA	0.1

Table 5.2: Parameters used in activation dispersal rule and learning rule.

relationships. Third, the map drawing of each feature is threshold dependent. When doing a visual comparison, some of the less prominent elements could be ignored. Further, the strongest elements appear on the map equally with other elements above the threshold. Therefore, no matter what threshold is chosen, there is a certain bias within the comparison.

The similarity measuring method is able to avoid the above limitations. This method is a quantitative measurement of the correlation of two features. It can indicate whether two feature are fully correlated, partially correlated, or not correlated at all. It is also quite simple to see the one-to-many feature correlations by listing the pairwise similarity values between all the relevant features. The similarity value is threshold independent, which means that it accounts for all of the involved elements, to a degree based on their activation levels. Therefore, the similarity value is a more precise measurement of the correlations of features. The disadvantage of this measurement is that it is less intuitive. Also the single value does not reflect the distribution of the cortical maps. So this method is only suitable in the analysis of the correlations of the feature maps.

## 5.3 Results

In this section, first the cortical feature maps in MI, including visual input map and motor output map, are described. Then the relationship between these input maps and output maps are investigated.

### 5.3.1 Cortical Map Formation in MI

Fig. 5.4 shows the MI output maps for upper arm extensor (E) and flexor (F), before and after training. The maps for other muscles show similar properties (See Appendix A.2.1 for a complete list of maps of all muscles). It is quite clear that before training, the MI elements that control the same muscle are randomly distributed throughout the MI layer, due to the random initialization of the connection weights between the MI layer and the lower motor neuron layer. After training, elements controlling each muscle have aggregated to form clusters. The size of the clusters varies, but single element clusters are unlikely. This is due to the ‘‘Mexican Hat’’ shape of the cortical activation pattern during training. The difference of the maps before and after training indicate

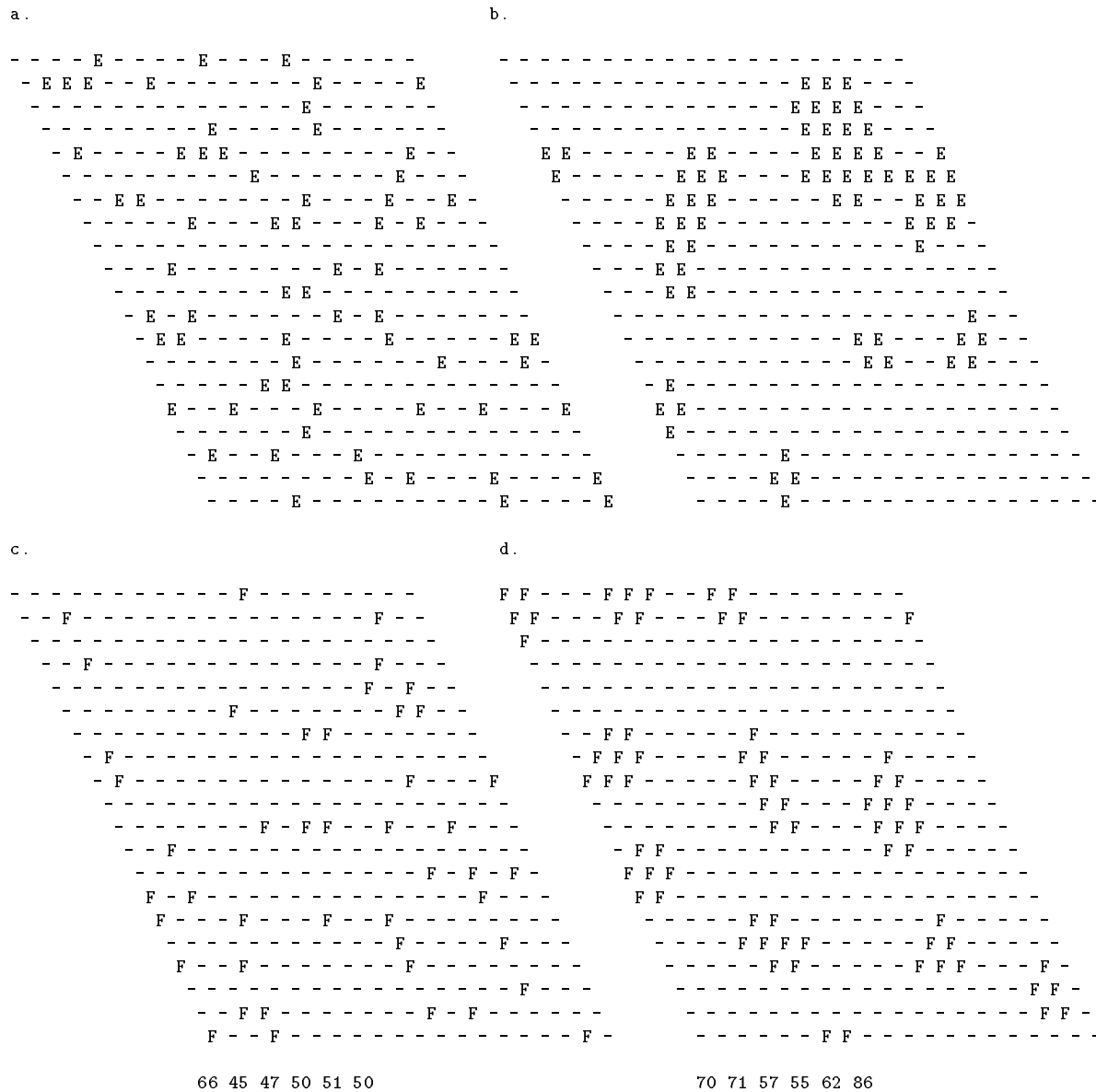


Figure 5.4: The MI output maps before (left) and after (right) training. Only maps for the upper arm extensor (E) and flexor (F) are shown (threshold=0.4).

that self-organization has occurred. Also examining the post-training maps in detail indicated that the maps of antagonist muscles are mutually exclusive (no overlap). The characteristics of these maps are similar to the motor output maps in the model with proprioceptive input only.

Fig. 5.5 shows the MI input maps from the visual input layer, before (left) and after (right) training. Only maps in the X dimension are shown. Maps in the other dimensions show similar characteristics (See Appendix A.2.2 for a complete list of maps of all dimensions). Each ‘X1’ in the map indicates that the MI element in that position is tuned to the stimulation of visual input element representing the negative range of the X dimension. Similarly, ‘X2’ and ‘X3’ represent elements tuned to the middle and positive range of the X dimension, respectively. Although it is less apparent in characterizing the nature of the maps before and after training, some changes can still be observed when examining these maps in detail. First, there is a tendency to form larger clusters during training. It can be seen that there are many single element clusters in the maps before training, while this is unlikely to occur after training. Quantitatively speaking, the total number of clusters in these three maps changed from 46 before training to 20 after training, a 57% decrease. The total number of tuned elements changed from 108 to 94 in the meantime, only a 13% decrease. This indicates that on average clusters are bigger after training, as is evident by visual inspection of Figure 5.5. In fact the average number of elements per cluster grows from 2.35 before training to 4.7 after training. The aggregation tendency is less apparent than we saw in the MI output maps because of the intracortical connections in MI, which has the tendency to form “Mexican Hat” shape of clusters even when the input connections are random before training. Second, it can be seen that there is a dramatic shift of the elements that are tuned to the stimulation of the same visual input element. There is little overlap between the same map before and after training. This type of reorganization during training is believed to be important in forming the correct input-output correlations, as will be seen in Section 5.3.2. Also the maps of X2 and X3 are almost identical (with similarity measuring value of 0.95) because during training the hand position could move to near the origin (shoulder) in the X dimension (X2) or to the back of the body (X3) only when upper arm extensor muscle is strongly contracted. As a result, the X2 and X3 maps all become correlated with E output maps, as will be seen in the next section.

### 5.3.2 Correlations Between the MI Input and Output Maps

In Section 5.3.1, the MI input and output maps have been shown to possess certain characteristics (aggregating, etc.) after training. These properties are quite similar to those reported in previous chapter. It is also interesting to study the relationships between the input maps and the output maps. Unlike the relationships between proprioceptive input maps and motor output maps, which are related to the same agonist/antagonist muscle groups, the correlation of input maps from the visual afferent pathway and the output maps to the muscle efferent pathway are more complicated.

Before training, there are no clear correlations between visual input and motor output features. All of the feature distributions are random. While this is less likely to be seen by visual comparison of feature maps, it is quite clear when using the similarity measuring method. Table 5.3 shows all of the similarity values between visual input and motor output features before training. In this table, each column represents the correlations of different visual input features to the same indicated motor output feature, while each row represents the correlations of different motor output features to the same visual input features. For example, the value 0.36 in column F and row X3 is the

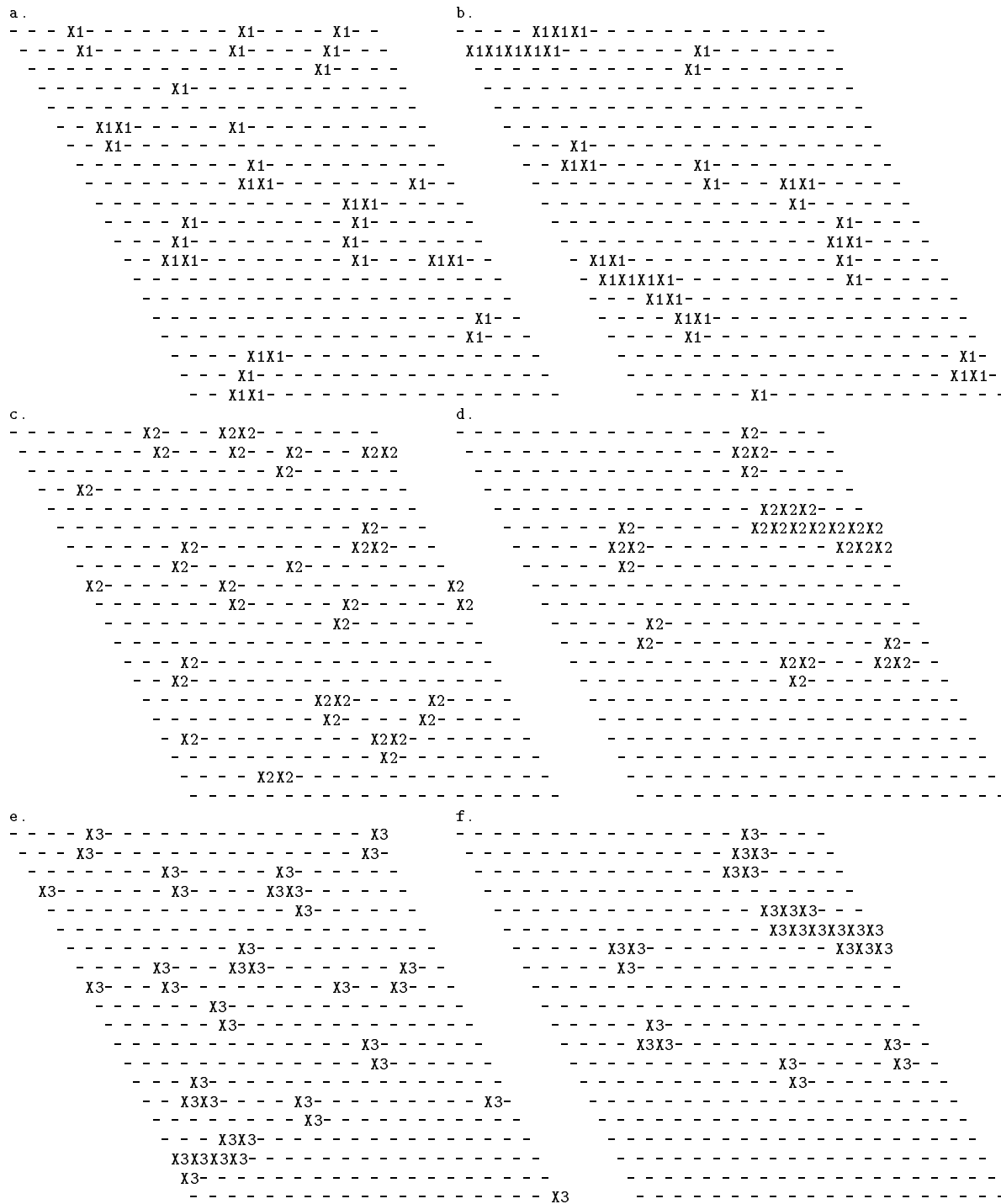


Figure 5.5: The MI input maps with respect to visual input (in the X dimension), before (left) and after (right) training. X1, X2 and X3 code the negative, middle and positive range in the X dimension.

	E	F	B	D	O	C
X1	0.38	0.37	0.37	0.35	0.37	0.37
X2	0.34	0.37	0.36	0.37	0.32	0.37
X3	0.35	0.36	0.39	0.34	0.41	0.37
Y1	0.39	0.40	0.37	0.41	0.38	0.39
Y2	0.41	0.35	0.39	0.40	0.37	0.38
Y3	0.33	0.34	0.35	0.35	0.37	0.36
Z1	0.35	0.37	0.40	0.37	0.39	0.38
Z2	0.38	0.36	0.37	0.36	0.37	0.34
Z3	0.37	0.39	0.38	0.37	0.37	0.35

Table 5.3: Similarity values between visual input and motor output features before training.

similarity value between the motor output feature F (upper arm flexor) and the visual input X3 (positive range of X). In this table, all of the similarity values are quite similar, ranging between 0.32 and 0.41. Since the distributions of features are initially quite random, no pair of features is either strongly correlated (close to 1.0) or strongly anticorrelated (close to 0.0).

After training, the MI visual input maps and motor output maps have both reorganized to form meaningful relationships, so that the model performs in a way that reflects the arm mechanism constraints. Unlike the correlations between the proprioceptive input maps and the motor output maps described in the previous chapter, the correlations between the visual input maps and the motor output maps do not have a clear one-to-one correspondence. Instead, each visual input feature can correlate with multiple motor output features, and *vice versa*. As a result, the maps of two correlated features are no longer fully aligned, as was seen earlier. There is a partial alignment of the maps. That is, the elements that represent both features are only partially overlapped, and the degree of overlap depend on how strongly the two features correlate. For example, Fig. 5.6 shows the two maps, E (upper arm extensor) and Y3 (positive Y axis), that have very strong correlations. As can be seen, there are quite a number of elements that are tuned to both features. Other strongly correlated pairs include (F,Y1), (B,Z3) and (D,Z1), all having similarity values of over 0.8. These strongly correlated pairs can be examined by visual comparison of their maps. For other pairs of correlated features, the partial alignments of the maps make it less apparent during visual comparison. It is therefore useful to use the quantitative approach, the similarity measuring method, to investigate the correlations of input and output features.

Table 5.4 shows the matrix of all the similarity values between visual input and motor output features after training. These values give a precise measurement of correlations of input and output features. It is apparent from the table that after training, the input and output feature maps have reorganized to form various degree of correlations between each other. The similarity values range from 0.01 to 0.89. Those values bigger than 0.41 (the biggest value in Table 5.3) are believed to represent pairs of features that are correlated. For a more intuitive display, Fig 5.7 shows the density plot of the table using values in Table 5.4. The following paragraphs will show how these feature correlations are consistent with the arm mechanism and constraints in details. Referring to Fig. 5.2 is helpful in locating the hand position in space.

These relationships can be explained as follows:

1. Hand movements in the X dimension are mostly affected by E (upper arm extensor) and



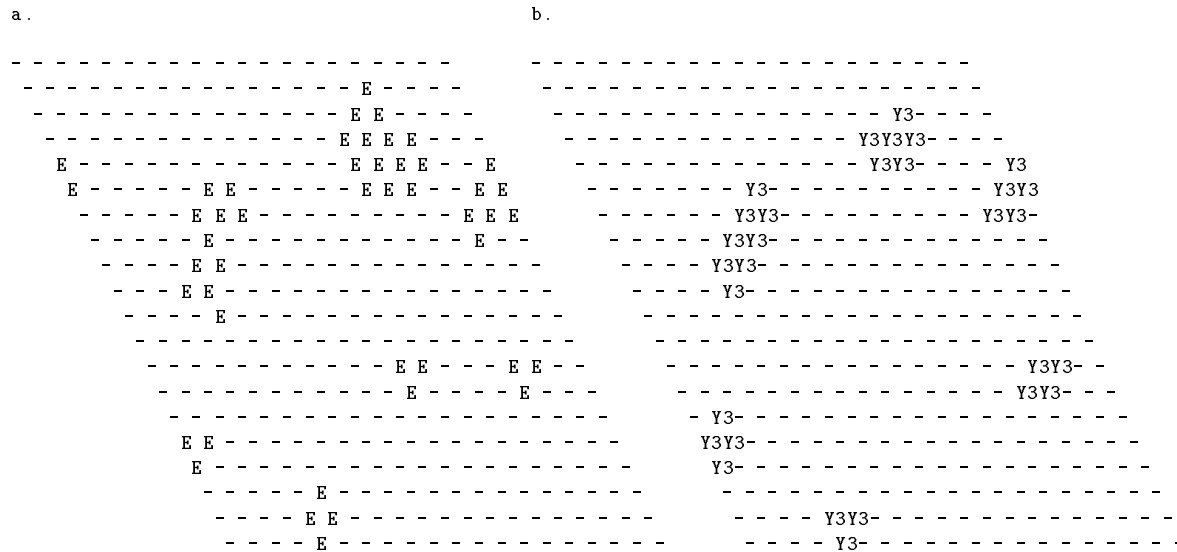


Figure 5.6: a. MI output map for upper arm extensor (E) after training (threshold=0.7); b. MI input map for negative range of the Y dimension (Y3) from visual input (threshold=0.3).

F (upper arm flexor). Feature X1 is correlated with F because when the upper arm flexor muscle is contracted, it is more likely to put the hand position in the negative range of the X axis, which is in front of the body. On the other hand, feature X2 and X3 are both correlated with E. When the upper arm extensor is contracted, the hand will move to the positive range of the X axis (back of the body) if the lower arm is fully extended. In case the lower arm is not fully extended, the final hand position would offset the displacement in the negative X direction and therefore it is quite likely that the hand position is compromised into the middle range of the X axis (around the Y-Z plane).

2. The hand movements in the Y dimension are affected by all the muscles. Feature Y1 is strongly correlated with F, indicating that the contraction of upper arm extensor is more likely to position the hand in the negative range of the Y axis (in front of the chest of a human body). On the other hand, feature Y3 is correlated with E and O. It means when upper arm extensor or the lower arm extensor is contracted, the arm is more likely to extend into the positive range of the Y axis (outside the right side of the body).

Feature Y2 is correlated with both B (upper arm abductor) and D (upper arm adductor) at the same time. This is quite interesting because not only one input feature is correlated with two output features, but also the two output features are antagonists, which tend to move the arm in opposite direction. On the other hand, this still reflects the arm mechanism and constraints, because when either the upper arm abductor or the upper arm adductor (but not both) are contracted, the elbow will be positioned very high or low near the X-Z plane. In both cases, the hand position will be moving within middle range of the Y axis in parallel with X-Z plane, no matter how the lower arm moves.

	E	F	B	D	O	C
X1	0.02	<b>0.59</b>	0.30	0.37	0.07	0.18
X2	<b>0.68</b>	0.01	0.03	0.02	0.13	0.02
X3	<b>0.58</b>	0.01	0.02	0.01	0.06	0.01
Y1	0.03	<b>0.89</b>	0.12	0.08	0.07	0.37
Y2	0.04	0.13	<b>0.46</b>	<b>0.63</b>	0.05	0.22
Y3	<b>0.87</b>	0.02	0.05	0.05	<b>0.67</b>	0.06
Z1	0.03	0.11	0.03	<b>0.88</b>	0.06	0.07
Z2	0.28	0.23	0.04	0.04	0.24	0.35
Z3	0.03	0.05	<b>0.84</b>	0.01	0.03	0.06

Table 5.4: Similarity values between visual input and motor output features after training. Those values that are bigger than the biggest values in Table 5.3 are in bold style.

3. The hand movements in the Z dimension are mostly affected by B (upper arm abductor) and D (upper arm adductor). Feature Z1 is strongly correlated with D, which is quite natural because contraction of upper arm adductor (folding the elbow toward the body) would lower the elbow position and therefore more likely put the hand position in the low range. For the same reason, Feature Z3 is correlated with B.

Feature of Z2 is not strongly correlated with any feature. However, Table 5.4 also shows that Z2 is actually weakly correlated with all feature other than B and D, with similarity values ranging from 0.24 to 0.35. It means that unless upper arm abductor or adductor are contracted to move the elbow up or down, it is all somewhat likely to put the hand position in the medium height.

4. The above paragraphs have shown the prominent motor output features that could effectively influence the hand position in each dimension. On the other hand, in Table 5.4, there are other less prominent similarity values which reflect some weak feature correlations. For example, feature X1 is also somewhat correlated with B and D, with similarity values of 0.30 and 0.37, respectively. That means, by clamping the upper arm vertically up or down, the hand would move along the semi-circle in X-Z plane, and therefore quite likely to be positioned in the negative range of the X axis, when the lower arm is in the middle range of the semi-circle. If the lower arm is fully folded or extended, then the X coordinate of the hand position will be close to 0, and therefore outside the negative of the X axis. This is why these two similarity values are less than that of  $(Y2, B)$  and  $(Y2, D)$ .

Another weakly correlated pair of features are Y1 and C, because the negative Y (in front of the chest) can be achieved only when upper arm flexor is somewhat contracted (to position the elbow near the X-Z plane ) and the lower arm flexor is contracted more than the lower arm extensor. However the lower arm flexor cannot be fully contracted (fully folding) because that will only maintain the hand position near the X-Z plane and therefore in the middle range of the Y axis. That is why the similarity value of  $(Y1, C)$  is not as big as  $(Y1, F)$ .

5. In Table 5.4, there are also similarity values that are close to 0, reflecting the anti-correlations between the features. Generally speaking, when one visual input feature is strongly correlated with a motor output feature of one muscle, it is usually anti-correlated with the output feature

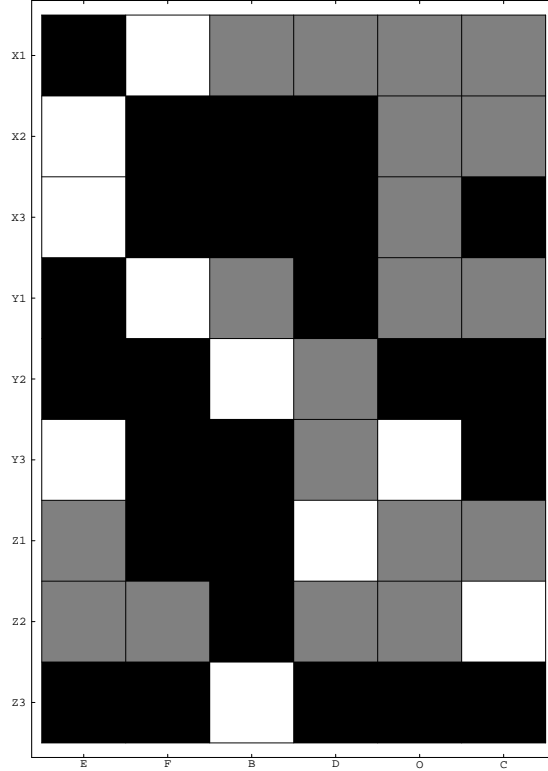


Figure 5.7: A Schematic Density Plot of Table 5.4. Those similarity values bigger than 0.4 (strongly correlated) are plotted as white blocks in their corresponding positions. The values smaller than 0.1 (anti-correlated) are plotted in black. Others (weak correlations) are plotted in grey.

of its antagonist muscle. The only exception is the pairs  $(Y2, B)$  and  $(Y2, D)$ , with the reason discussed in above paragraph. Another rule of thumb is that when a motor output feature is correlated with one visual input feature in one extreme of a dimension (not middle range), it is usually anti-correlated with the visual input feature in the other extreme of the same dimension.

In summary, the above analysis shows that after training, the cortical input and output maps have reorganized to reflect the arm mechanisms. It should be noted that these are qualitative analyses based on quantitative measurements. The values in Table 5.4 not only reflect the qualitative properties shown above, but also reflect how strongly they support these properties. Due to the randomness of the initial connection strength and the training patterns, the similarity values in Table 5.4 may not be the same in different simulations. I have run several simulations with different initial connection strengths and training patterns, and all of the properties described above still hold. Table 5.5 shows the similarity values in four other simulations with different initial weights and training patterns as examples.

	E	F	B	D	O	C		E	F	B	D	O	C
X1	0.02	<b>0.68</b>	0.35	0.20	0.06	0.20	X1	0.03	<b>0.52</b>	0.40	0.41	0.09	0.21
X2	<b>0.75</b>	0.01	0.02	0.02	0.18	0.03	X2	<b>0.70</b>	0.00	0.01	0.01	0.11	0.04
X3	<b>0.61</b>	0.01	0.01	0.01	0.04	0.01	X3	<b>0.60</b>	0.00	0.01	0.01	0.03	0.01
Y1	0.02	<b>0.74</b>	0.08	0.07	0.05	0.26	Y1	0.03	<b>0.82</b>	0.11	0.08	0.11	0.37
Y2	0.04	0.15	<b>0.53</b>	<b>0.47</b>	0.05	0.22	Y2	0.04	0.13	<b>0.48</b>	<b>0.47</b>	0.05	0.24
Y3	<b>0.77</b>	0.04	0.05	0.06	<b>0.76</b>	0.07	Y3	<b>0.86</b>	0.01	0.05	0.03	<b>0.66</b>	0.05
Z1	0.03	0.10	0.04	<b>0.88</b>	0.06	0.08	Z1	0.04	0.09	0.03	<b>0.81</b>	0.06	0.11
Z2	0.31	0.26	0.03	0.03	0.24	0.34	Z2	0.28	0.31	0.04	0.04	0.22	0.41
Z3	0.02	0.11	<b>0.88</b>	0.02	0.05	0.08	Z3	0.03	0.08	<b>0.83</b>	0.02	0.06	0.08
	E	F	B	D	O	C		E	F	B	D	O	C
X1	0.03	<b>0.55</b>	0.33	<b>0.42</b>	0.08	0.25	X1	0.02	<b>0.64</b>	0.31	0.32	0.09	0.20
X2	<b>0.82</b>	0.01	0.04	0.02	0.32	0.07	X2	<b>0.81</b>	0.01	0.03	0.02	0.28	0.03
X3	<b>0.63</b>	0.00	0.03	0.01	0.10	0.02	X3	<b>0.62</b>	0.00	0.01	0.02	0.11	0.01
Y1	0.03	<b>0.74</b>	0.08	0.08	0.06	0.38	Y1	0.02	<b>0.79</b>	0.07	0.06	0.08	0.30
Y2	0.04	0.12	<b>0.45</b>	<b>0.46</b>	0.04	0.23	Y2	0.03	0.12	<b>0.60</b>	<b>0.46</b>	0.04	0.17
Y3	<b>0.87</b>	0.04	0.05	0.05	<b>0.75</b>	0.12	Y3	<b>0.83</b>	0.02	0.07	0.05	<b>0.76</b>	0.07
Z1	0.03	0.09	0.03	<b>0.84</b>	0.07	0.07	Z1	0.02	0.09	0.02	<b>0.87</b>	0.05	0.05
Z2	0.28	0.33	0.05	0.04	0.18	0.34	Z2	0.29	<b>0.46</b>	0.06	0.07	0.31	<b>0.5</b>
Z3	0.04	0.07	<b>0.89</b>	0.02	0.06	0.06	Z3	0.02	0.08	<b>0.85</b>	0.02	0.05	0.08

Table 5.5: Similarity values between visual input and motor output features after training in four other simulations with different initial weights and training patterns. Those values that are bigger than the biggest values in Table 5.3 are in bold style.

## 5.4 Discussion

In the previous section, it was shown that the motor control model with visual inputs alone can form feature maps in the motor cortex layer during unsupervised learning. These maps and their relationships are important in achieving consistent control of the arm movement. The results indicate that, by supplying random activation patterns alone in motor cortex, the initial random cortical connections are able to self-organize to recognize the characteristics of arm mechanisms and form meaningful input output relationships. This is in contrast to most previous models, which are based on error minimizing instead of cortical maps. These models serve the purpose of certain tasks (such as reaching), and do not concern about internal representation (maps) of the outside side world.

In previous motor control model with proprioceptive input pathways (Chapter 4), we found that an input feature and an output feature that are temporally correlated with each other have their maps aligned. This is due to the one-to-one mapping of the input and output features. In the motor control model with visual inputs, the input-output relationships become more complicated. In this model, each input feature may be correlated with multiple output features, and *vice versa*. In this situation, the feature correlations could not be as strong as those in proprioceptive model. Assume an input feature A is correlated with output features B and C. As long as the maps of B and C are different, A's map cannot be fully aligned with B's and C's at the same time. So each input feature can only correlate strongly (with complete map alignment) with at most one output feature. In most cases, as we have seen, one feature is usually correlated with multiple features, with different correlation strengths. The quantitative measurement of the feature correlation provide us

with a way of studying these complicated input output relationships. *The similarity value between two features reflects the likelihood (or probability) that the two features are present simultaneous. The amount of temporal correlations of external events is represented internally via the degree of co-activation of cortical elements, i.e. as spatial correlations.* When two external events are closely associated (e.g., the output signal to contract a muscle and the input signal of increased tension of the same muscle), the cortical elements representing these two features are largely the same set. On the contrary, when two external event are mutually exclusive (e.g., the lengthening signals of both an arm muscle and its antagonist), the internal activation representing the two features become anti-correlated. That is, there are not likely to be any cortical elements responding strongly to both features. When two external events are weakly associated, there is a certain amount of overlap between the cortical elements tuned to both features.

Our current understanding of the coding of visual input information received by biological motor cortex is quite limited beyond primary visual cortex. Anatomical studies show that there is no direct neural projection from visual cortex to motor cortex: primary motor cortex receives visual information via secondary visual and other association areas [Felleman & Essen, 1991; Asanuma, 1989]. The coding of visual information received by MI is not known at present. However, MI does receive visual information coded in some form [Johnson, 1992]. Our model is a simple design attempting to incorporate the visual information. On the other hand, this model can be viewed in a more general framework: as a study of the one-to-many partially correlated feature associations. Basically, our brains can be viewed, from the computational point of view, as a multi-layer network mapping input sensory signals into output control signals. In each specific layer, the neuron elements responding to an input feature provide correct output feature(s) to the next layer. If each layer is doing a simple one-to-one feature mapping, then the entire brain functionality would be greatly limited. It is apparent that the one-to-many partially correlated feature mapping is important for the brain to exhibit versatile input output relationships and capabilities. By studying the relationships of this type of input and output information, we can have a better understanding about the internal representation and feature correlation of the brain.

## Chapter 6

# Motor Control Model with Combined Proprioceptive and Visual Inputs

In previous chapters, we have studied the motor control model based on proprioceptive input in isolation and visual input in isolation. In this chapter, another variation of the motor control model is studied. This version combines the models in the previous chapters, i.e., it is a model with both proprioceptive and visual inputs.

Fig. 6.1 shows a schematic diagram of the model considered here. In this model, the motor cortex layer (MI) and the proprioceptive cortex layer (PI) are 20 by 20 two-dimensional, hexagonally tessellated layers, with each element connected to its six neighbors. Each element in PI is connected to its corresponding element in MI and the surrounding MI elements up through a radius of four, forming a coarse topographic ordering. Each of twelve elements in the proprioceptive input layer, coding the length and tension information of six muscles, is fully connected with the PI layer. And each element in the visual input layer, coding hand position, is fully connected to the MI layer. The MI layer is also fully connected to the lower motor neuron layer, which has six elements representing the average activation levels of six muscles. The transformation from lower motor neuron activation to proprioceptive input information and the hand position is based on the mechanism of the model arm. The coding of visual information is the same as in the previous chapter.

### 6.1 Experimental Methods

The training method is similar to that used in previous chapters. All the interlayer weights were randomly initialized, while the intralayer connections in PI and MI layer are of fixed strength. Training was done by applying a patch of activation (of radius 1) at randomly selected MI regions. The activation then spread to the lower motor neuron layer. For a given activation pattern in the lower motor neuron layer, the length and tension information of muscles could be calculated, as could the spatial information of hand position, according to the arm mechanism. The muscle length and tension information was then supplied to the proprioceptive input layer. Simultaneously the hand position information was supplied to the visual input layer, according to the visual coding mechanism described in the previous chapter. The proprioceptive input layer sends activation to the PI layer, which subsequently sent activation to MI. The visual input layer sent activation directly to MI. The MI layer received feedback information from both the PI layer and the visual input layer and the combined activation jointly determined the activation in the MI layer and thus muscle activations. This formed a closed-loop system, in which no element activations were clamped in each

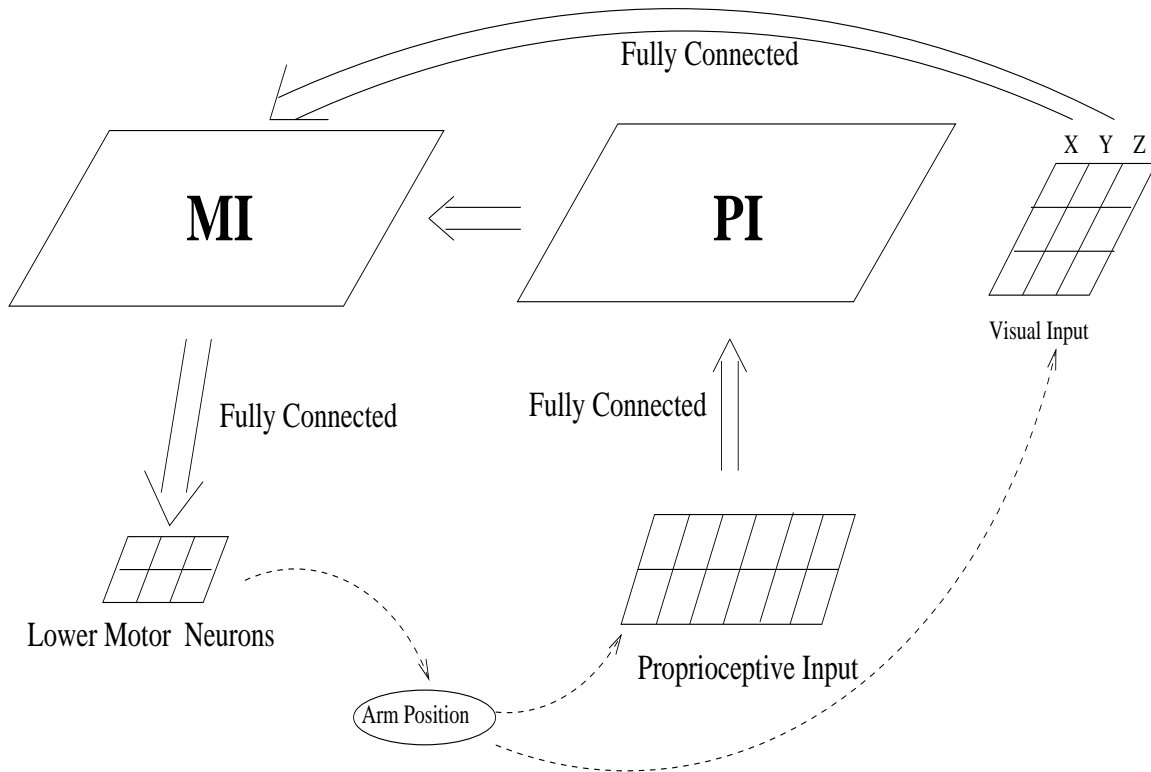


Figure 6.1: The motor control system with both proprioceptive and visual inputs. The MI elements send activation to lower motor neurons, which direct arm movements. Information about the arm configuration and hand position is received by both the proprioceptive and visual input layers. Both input layers then supply this information to the MI layer. Feedback to the MI layer influences the MI output and thus forms a closed-loop.

layer. The spread of activation was governed by the activation rule using competitive distribution of activation (Equation 3.8 and 3.9). After sufficient time steps (120 were used in this experiment), the stabilized activation levels were achieved in all of the layers. Then learning was conducted by applying the competitive learning rule to all of the inter-layer connections (Equation 3.10 and 3.11). The model was trained with 2000 learning cycles before the maps were examined.

The parameter values used in Equations 3.8-3.11 of the model in this experiment are summarized in Table 6.1 and 6.2. The learning threshold,  $\alpha$ , is 0 in all layers. The parameters used in this experiment are similar to those in the previous experiments. Any small change to any parameter would not change the qualitative characteristics of the results discussed in following section.

After training, all of the cortical input and output maps, along with their relationships, were examined. These maps include: PI input maps from proprioceptive inputs; MI input maps from proprioceptive inputs; MI input maps from visual inputs; and MI output maps to lower motor neurons. The measuring methods with these maps were similar to those described in previous chapters. Basically, the output maps were determined by stimulating the cortical elements and measuring the activities in the output layer. The input maps were measured by stimulating the elements in the (proprioceptive or visual) input layer one by one and measuring the corresponding

Parameters	PI layer	MI layer	Motor Neurons
$c_s$	-4.0	-2.0	-0.2
M	5.0	3.0	1.2

Table 6.1: Parameters used in activation update rule.

Parameters	Arm to PI	PI to PI	PI to MI	MI to MI	MI to Motor	Visual to MI
q	0.04	0.0001	0.0001	0.0001	0.0001	0.0001
p	1	1	1	1	2	1
$c_p$	0.8	0.7	0.3	0.4	0.005	5.0
$\eta$	0.2	NA	0.2	NA	0.1	0.1

Table 6.2: Parameters used in activation dispersal rule and learning rule.

cortical activities. In both cases, the elements that were activated above a certain threshold were used during map drawing. Also, in the measurement of input maps, the corresponding input layer was stimulated and then clamped, and cortical activations were measured after activation stabilization was achieved. In order to study the relationships between the input maps and output maps, the similarity measuring method of Equation 4.1 was also used to quantitatively measure the correlations of input and output features.

## 6.2 Results

### 6.2.1 PI Input Maps with Respect to the Proprioceptive Inputs

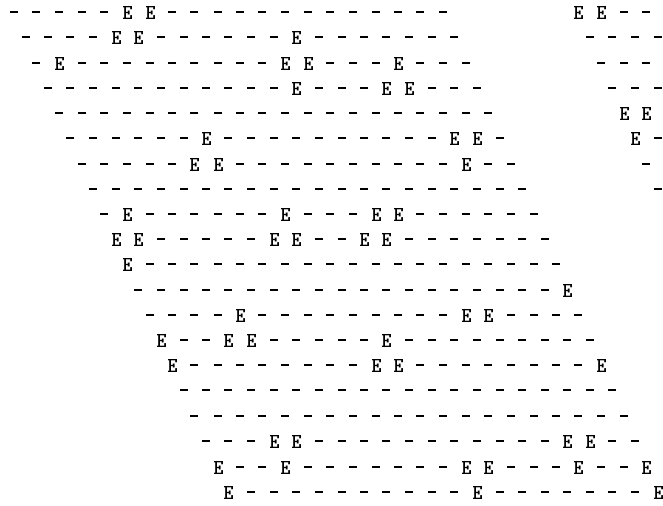
Fig. 6.2 shows the PI input maps with respect to proprioceptive inputs, before (left) and after (right) training. Only the maps with respect to the length input of the upper arm extensor and flexor muscles are shown. Length and tension maps for other muscles show similar characteristics (See Appendix A.3.1 for a complete list of maps of all muscles). From Fig. 6.2, it is clear that activation clusters formed both before and after training due to the “Mexican Hat” activation patterns induced by intracortical connections. However, the clusters were more regularly arranged after training. This indicated the self-organization of the proprioceptive cortical maps during training. A similar self-organization occurred in the motor control model with proprioceptive input only, as described in Chapter 4. No qualitative difference between the model with proprioceptive input alone and the current model was observed.

The self-organization of feature maps in the proprioceptive cortex can be more clearly seen when examining the relationships between these input maps. As in the motor control model with proprioceptive input only, *the length map of a particular muscle matches well with the tension map of its antagonist muscle*. For example, Fig. 6.3 shows that the length map of the upper arm extensor matches the tension map of upper arm flexor; and the length map of the upper arm flexor matches the tension map of upper arm extensor. For other pairs of muscles, there are similar relationships. This means that this previously described map alignment property was preserved after visual input was added into the model.

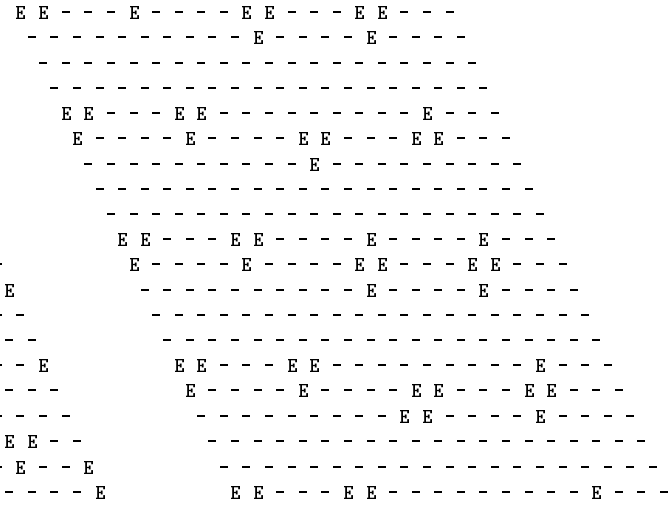
Table 6.3 shows two tables of similarity values, before (left) and after (right) training, respec-



a.



b.



c.



d.

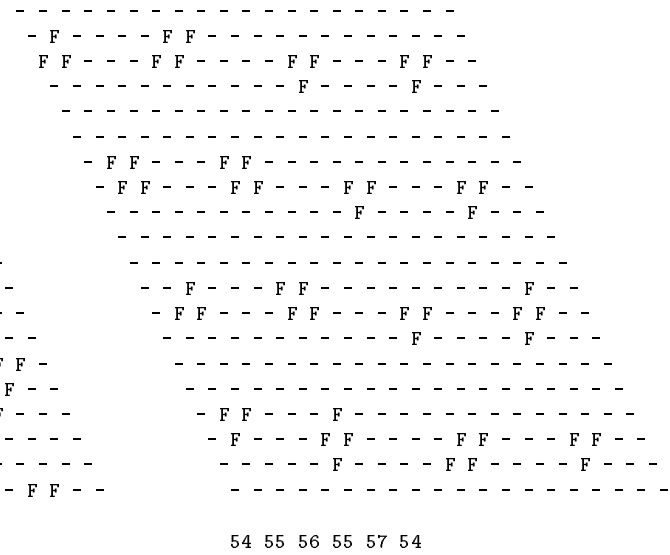


Figure 6.2: The PI input maps before (left) and after (right) training. Only length maps for the upper arm extensor (E) and flexor (F) are shown (threshold=0.2).

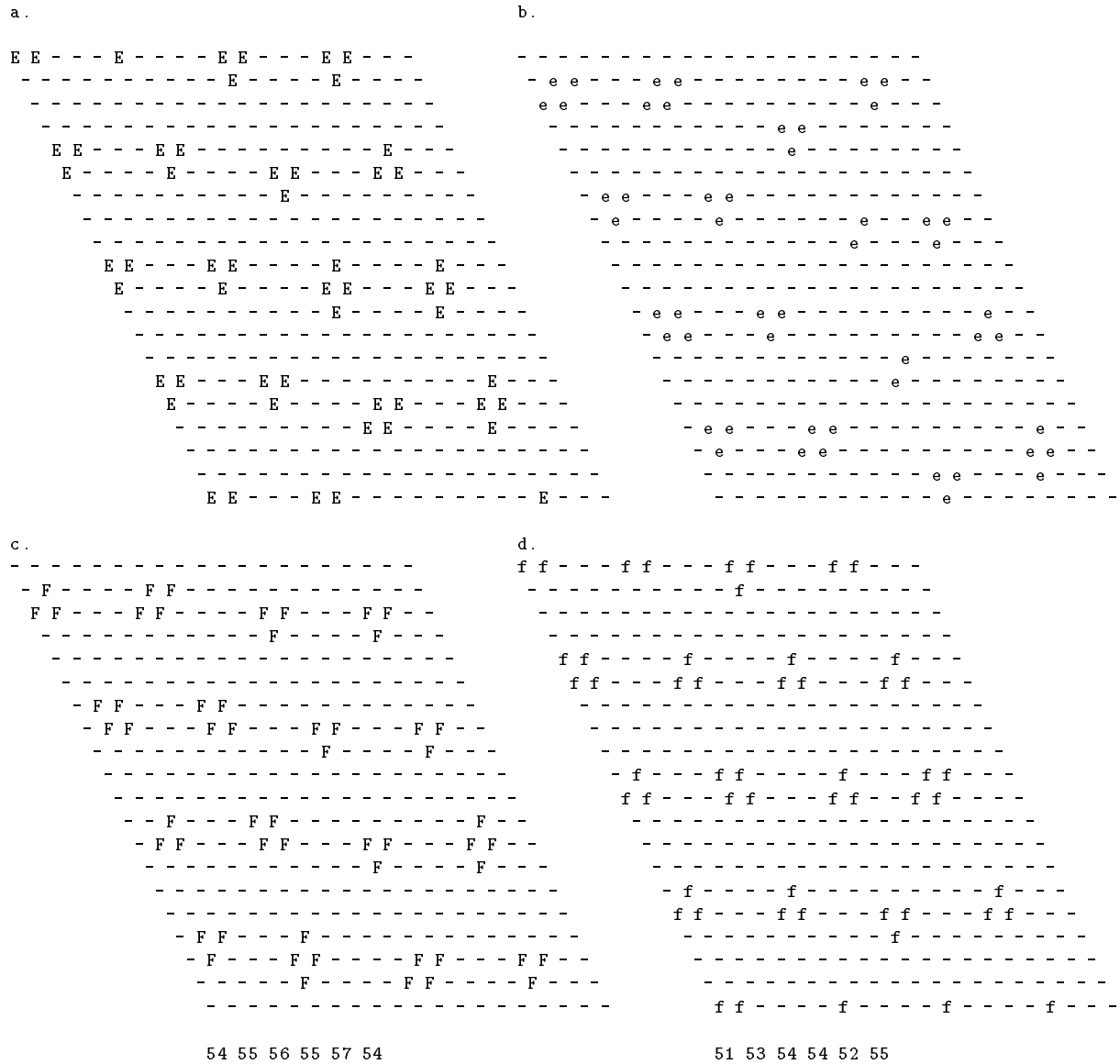


Figure 6.3: Comparison of PI input maps with respect to length (left) and tension (right) after training. Only maps for the upper arm extensor (E,e) and flexor (F,f) are shown (threshold=0.2).

	E	F	B	D	O	C		E	F	B	D	O	C
e	0.11	0.10	0.26	0.21	0.17	0.15	e	0.01	<b>0.84</b>	0.17	0.08	0.11	0.06
f	0.32	0.16	0.11	0.17	0.17	0.17	f	<b>0.91</b>	0.00	0.14	0.01	0.02	0.42
b	0.12	0.15	0.20	0.20	0.05	0.37	b	0.01	0.11	0.00	<b>0.96</b>	0.02	0.13
d	0.12	0.19	0.07	0.16	0.35	0.20	d	0.01	0.05	<b>0.80</b>	0.00	0.24	0.00
o	0.13	0.17	0.14	0.15	0.18	0.14	o	0.56	0.06	0.17	0.04	0.02	<b>0.85</b>
c	0.12	0.19	0.10	0.14	0.11	0.29	c	0.04	0.28	0.26	0.02	<b>0.94</b>	0.00

Table 6.3: Similarity values between length and tension input features in proprioceptive cortex before (left) and after (right) training. Those values that are bigger than 0.7 are in bold style.

tively. Each row of the table is for a particular muscle’s tension feature, while each column is for a particular muscle’s length feature . Each value in the table represents the similarity measurement between the length feature of the corresponding column and the tension feature of the corresponding row. For example, the value 0.11 in the upper left corner of the table represents the similarity measurement between the length and tension features of the upper arm extensor before training. Before training, similarity values ranges from 0.05 to 0.37, due to the activations produced by the initial random weights. After training, some pairs of features exhibited strong correlations. Those pairs having similarity values bigger than 0.8 are the length and tension features of mutually antagonist muscles (bold type in Table 6.3). This is clear evidence that the relationships within proprioceptive cortical input maps in the original model still holds in this variation of the model, and that this kind of relationship is clearly the result of training.

### 6.2.2 MI Proprioceptive Input Maps

MI input maps with respect to the proprioceptive input also formed after training. After training, the activation clusters were slightly more uniform in size and regularly arranged. Particularly, the relationship between the length and tension input maps after training clearly reflected the arm mechanism and constraints. Fig. 6.4 shows the MI length (left) and tension (right) input maps of the upper arm extensor (E,e) and flexor (F,f) after training. It is quite clear that the length map of the upper arm extensor (E) matches the tension map of upper arm flexor (f); and the length map of the upper arm flexor (F) matches the tension map of upper arm extensor (e). This relationship formed because during training, the contraction of one muscle usually increased its tension and decreased its length (and therefore increased its antagonist’s length). This relationship was true for both PI and MI layer, indicating that although some transformation occurred in the proprioceptive input maps from PI to MI, resulting in complete different maps in two cortices, same internal correlations were still preserved.

The relationships between the length and tension input maps were also examined with the similarity measuring method. Table 6.4 shows the similarity values between length and tension input features in motor cortex before (left) and after (right) training, respectively. Before training, the similarity values ranged randomly from 0.26 to 0.63. After training, those pairs that are strongly correlated (having similarity values bigger than 0.8) happened to be the length and tension features of mutually antagonist muscles.

In summary, the relationships in the proprioceptive input patterns were captured by both PI and MI during training, and were reflected in the post-training proprioceptive input maps in both

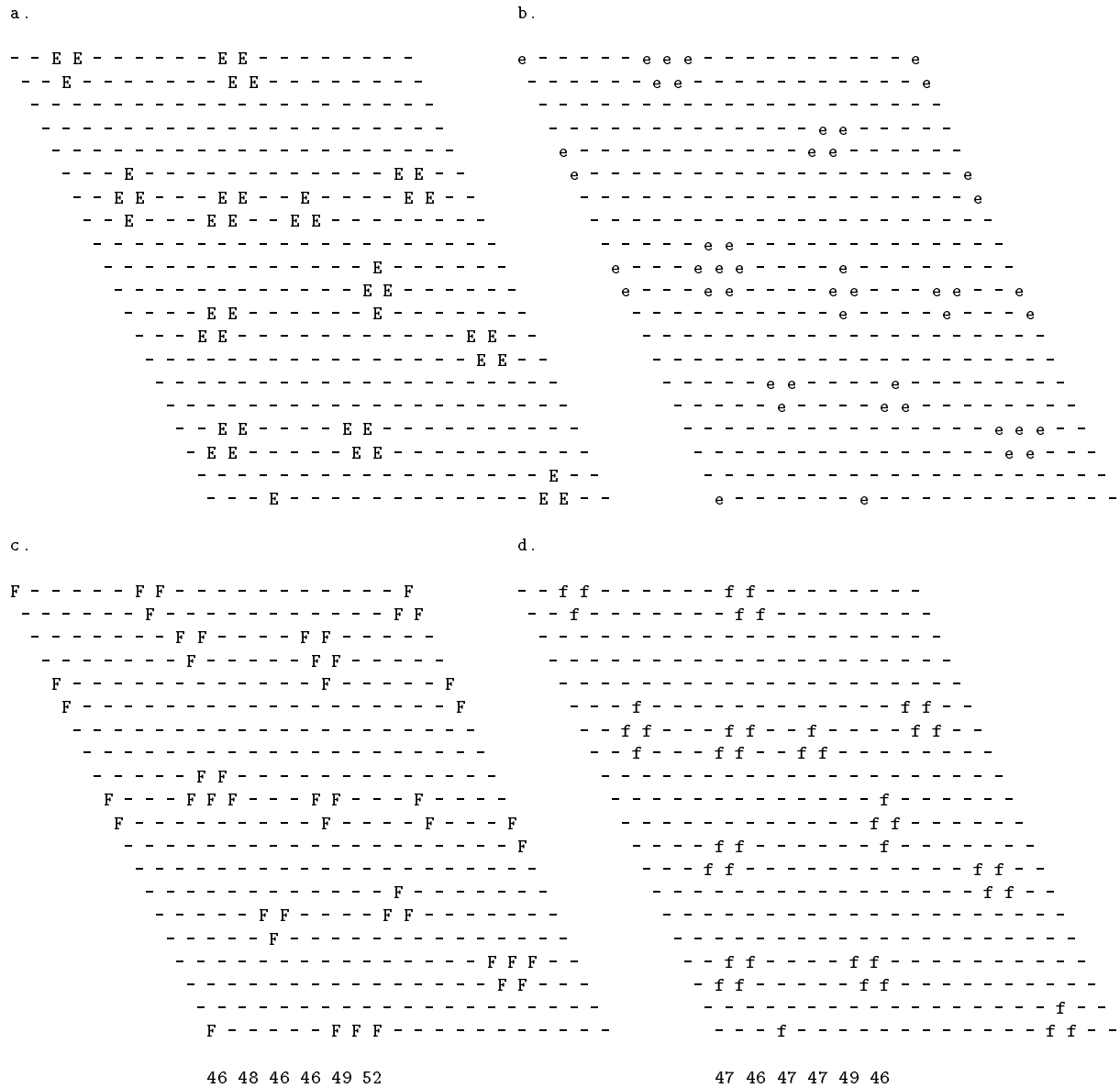


Figure 6.4: Comparison of MI input maps with respect to length (left) and tension (right) after training. Only maps for the upper arm extensor (E,e) and flexor (F,f) are shown (threshold=0.4).

	E	F	B	D	O	C		E	F	B	D	O	C
e	0.42	0.42	0.47	0.50	0.46	0.48	e	0.00	<b>0.85</b>	0.04	0.00	0.04	0.11
f	0.63	0.50	0.43	0.59	0.41	0.46	f	<b>0.99</b>	0.00	0.05	0.00	0.03	0.27
b	0.45	0.41	0.48	0.48	0.26	0.52	b	0.00	0.02	0.02	<b>0.99</b>	0.02	0.20
d	0.37	0.43	0.33	0.32	0.47	0.49	d	0.02	0.01	<b>0.94</b>	0.03	0.05	0.01
o	0.40	0.39	0.37	0.42	0.39	0.39	o	0.50	0.05	0.03	0.00	0.01	<b>0.78</b>
c	0.52	0.48	0.41	0.48	0.35	0.57	c	0.05	0.20	0.10	0.01	<b>0.97</b>	0.00

Table 6.4: Similarity values between length and tension input features in motor cortex before (left) and after (right) training. Those values that are bigger than 0.7 are in bold style.

cortical layers. These experimental results are similar to those reported in the original motor control model without visual input, as described in Chapter 4. Thus, adding visual inputs to the model had little impact on the proprioceptive maps that formed in PI and MI.

### 6.2.3 MI Output Maps and Their Relation to Proprioceptive Input Maps

The output map of the MI layer to the lower motor neurons were also examined with this model. Fig. 6.5 shows the MI output maps of upper arm extensor muscle before (left) and after (right) training (See Appendix A.3.3 for a complete list of maps of all muscles). Before training, the output maps exhibited a random arrangement. After training, the elements representing the same feature tended to form clusters that are uniform in size and arrangement. The maps of other muscles showed similar characteristics. This indicates the self-organization of cortical feature maps during training, and this kind of self-organization also occurred in the models with proprioceptive or visual input only, as described in previous chapters.

The effect of self-organizing feature maps in MI was more clearly seen when the relationships between the MI output maps and the MI proprioceptive input maps were studied. Fig. 6.6 shows the comparison of the MI length and tension input maps of upper arm extensor and the MI output maps of upper arm extensor and flexor after training. It is clear that the length map of upper arm extensor (Fig. 6.6a) matches the output map of upper arm flexor (Fig. 6.6d); while the tension map of upper arm extensor (Fig. 6.6b) matches the output of upper arm extensor (Fig. 6.6c). Examining maps of other muscles revealed a general rule: *the length input map of one muscle matches the motor output map of its antagonist muscle; and the tension input map of one muscle matches the motor output map of itself*. This rule is also clearly seen with the similarity measuring method. Table 6.5 summarized the similarity values between MI input and output features before (left) and after (right) training. In each column of the table, the values correspond to the same output feature, represented by a capital letter on top of the column. In each row of the table, the values correspond to the same length or tension input feature, represented by a capital or lower case letter, respectively, on the left hand side of the row. For example, in the table on the left hand side, the value 0.48 in the row of 'E' and column of 'B' is the similarity value between the length input feature of upper arm extensor (E) and the motor output feature of upper arm abductor (B) before training. In the same table, the value 0.45 in the row of 'e' and the column of 'E' is the similarity value between the tension input feature of upper arm extensor (e) and the motor output feature of upper arm extensor (E). From Table 6.5, one can see that before training the similarity values ranged randomly from 0.44 to 0.54, showing no strong correlations. After training, the similarity

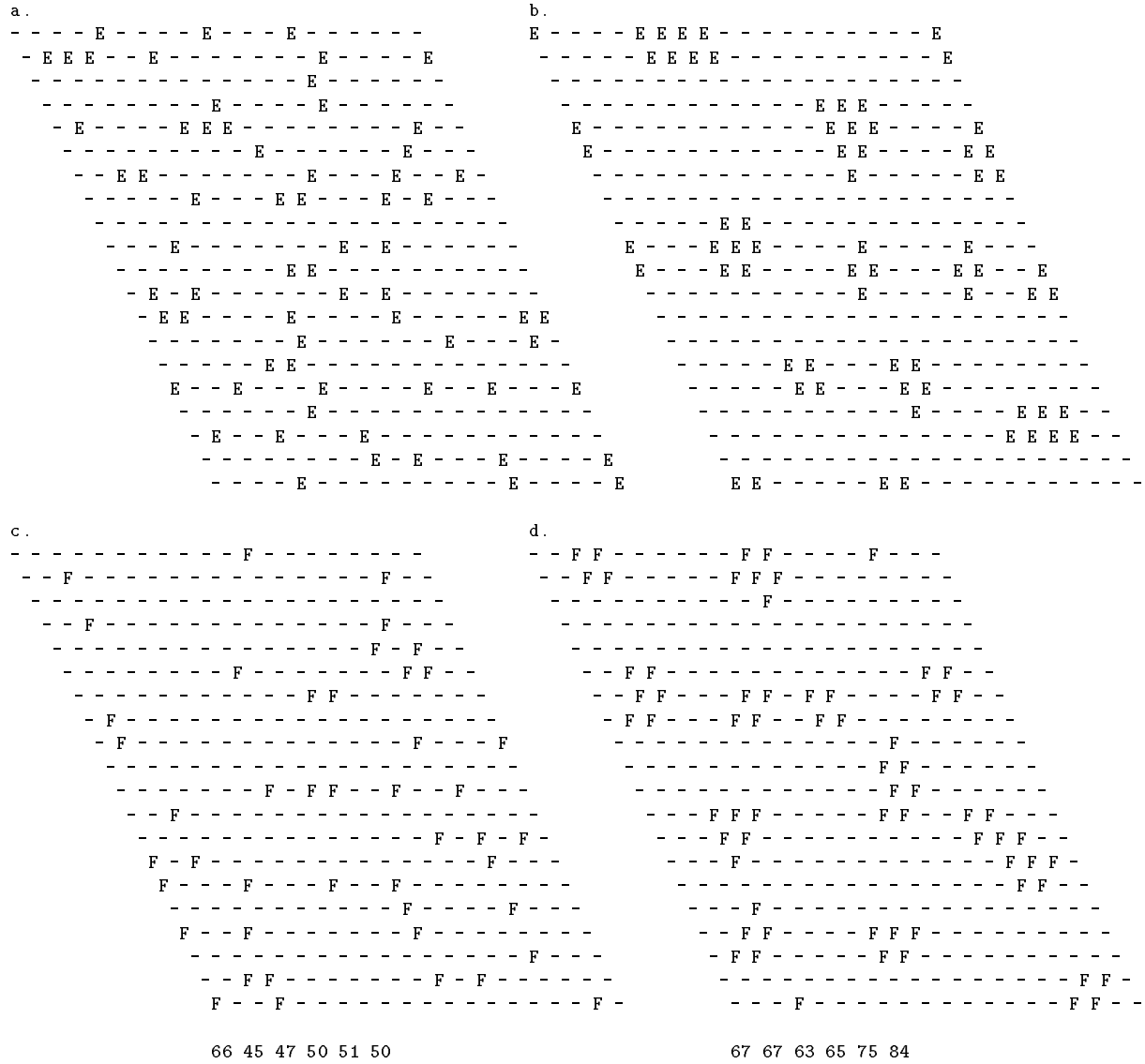


Figure 6.5: The MI output maps before (left) and after (right) training. Only maps for the upper arm extensor (E,e) are shown (threshold=0.4).

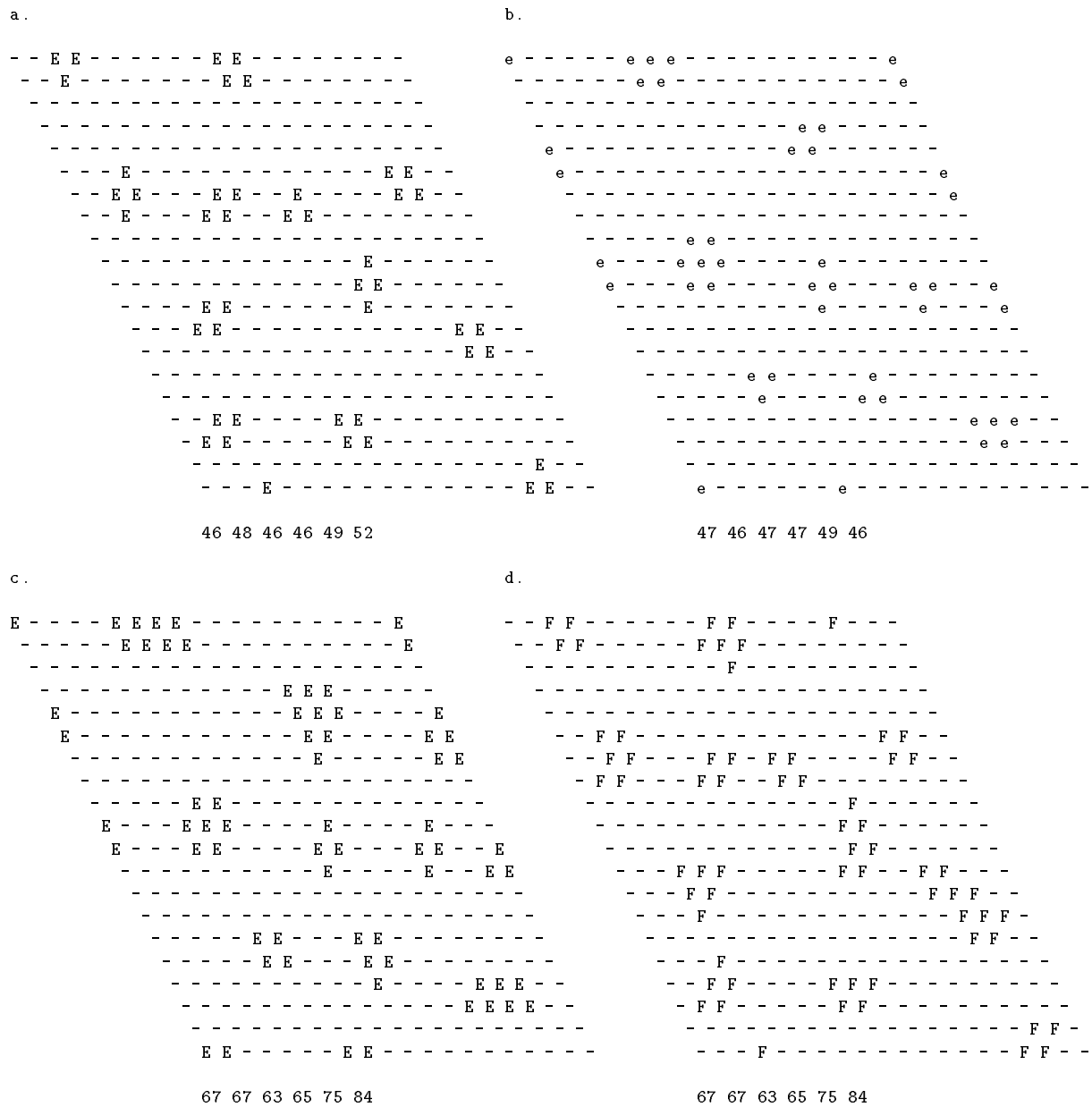


Figure 6.6: The comparison of the MI input maps with respect to the length (a) and tension (b) of upper arm extensor, and the MI output maps of upper arm extensor (c) and flexor (d) after training (threshold=0.4).

	E	F	B	D	O	C		E	F	B	D	O	C
E	0.49	0.49	0.48	0.49	0.47	0.47	E	0.05	<b>0.96</b>	0.11	0.10	0.31	0.18
F	0.48	0.47	0.48	0.48	0.51	0.45	F	<b>0.81</b>	0.02	0.08	0.07	0.29	0.28
B	0.45	0.47	0.48	0.46	0.47	0.47	B	0.13	0.14	0.05	<b>0.94</b>	0.16	0.23
D	0.47	0.48	0.51	0.49	0.49	0.51	D	0.08	0.05	<b>0.97</b>	0.07	0.05	0.05
O	0.50	0.48	0.49	0.46	0.47	0.48	O	0.12	0.09	0.08	0.11	0.08	<b>0.89</b>
C	0.50	0.50	0.53	0.51	0.52	0.54	C	0.26	0.31	0.29	0.08	<b>0.74</b>	0.07
e	0.45	0.46	0.49	0.46	0.47	0.46	e	<b>0.94</b>	0.03	0.06	0.08	0.43	0.13
f	0.49	0.48	0.50	0.51	0.49	0.46	f	0.05	<b>0.95</b>	0.10	0.09	0.30	0.17
b	0.44	0.47	0.44	0.45	0.47	0.48	b	0.08	0.05	<b>0.96</b>	0.07	0.05	0.04
d	0.51	0.51	0.50	0.49	0.52	0.51	d	0.10	0.08	0.06	<b>0.94</b>	0.13	0.14
o	0.45	0.47	0.49	0.45	0.44	0.47	o	0.26	0.52	0.08	0.08	<b>0.82</b>	0.11
c	0.49	0.47	0.48	0.49	0.49	0.46	c	0.10	0.10	0.06	0.09	0.06	<b>0.88</b>

Table 6.5: Similarity values between length and tension input features in motor cortex before (left) and after (right) training. Those values that are bigger than 0.7 are in bold style.

values changed dramatically to reflect the correlations of the self-organized feature maps. The upper half of the table on the right hand side illustrates the correlations between the length input feature and the motor output feature after training. One can see that pairs of antagonist muscles have strong correlations, with similarity values from 0.74 to 0.97. The lower half of the table on the right hand side illustrates the correlations between the tension input feature and the motor output feature after training. It is quite clear that all of the large similarity values (from 0.82 to 0.96) are on the diagonal line, indicating that the tension input feature and the motor output feature of the same muscle are strongly correlated. All of these properties are natural results of the training. The same properties occurred in the motor control model with proprioceptive input only.

#### 6.2.4 MI Visual Input Maps and Their Relation to MI Output Maps

The formation of proprioceptive input maps and the motor output maps, as well as their relationships to each other, are similar to the motor control model with proprioceptive input only. In this section, the MI visual input maps are summarized, as well as their relationships with the motor output maps. It is particularly interesting to compare the results in this model with the model in Chapter 5, which has visual input only.

Fig. 6.7 shows MI visual input maps before (left) and after (right) training. Only the visual input maps for the X dimension are shown. The maps for other dimensions show similar characteristics (See Appendix A.3.4 for a complete list of maps of all dimensions). The representation of the visual input maps is the same as that described in Chapter 5. When comparing the visual input maps before and after training, it is quite clear that the maps before training are quite different from those after training, indicating some kind of self-organization during training. Although the maps both before and after training formed activation clusters, the clusters after training were more uniform in size and shape. Statistically, before training the average number of elements in each cluster in the three X-dimensional input maps was 3.04, with standard deviation of 1.41. After training average size of clusters increased to 3.46, with standard deviation decreased to 0.85. So the activation clusters after training became bigger and varied less in size.



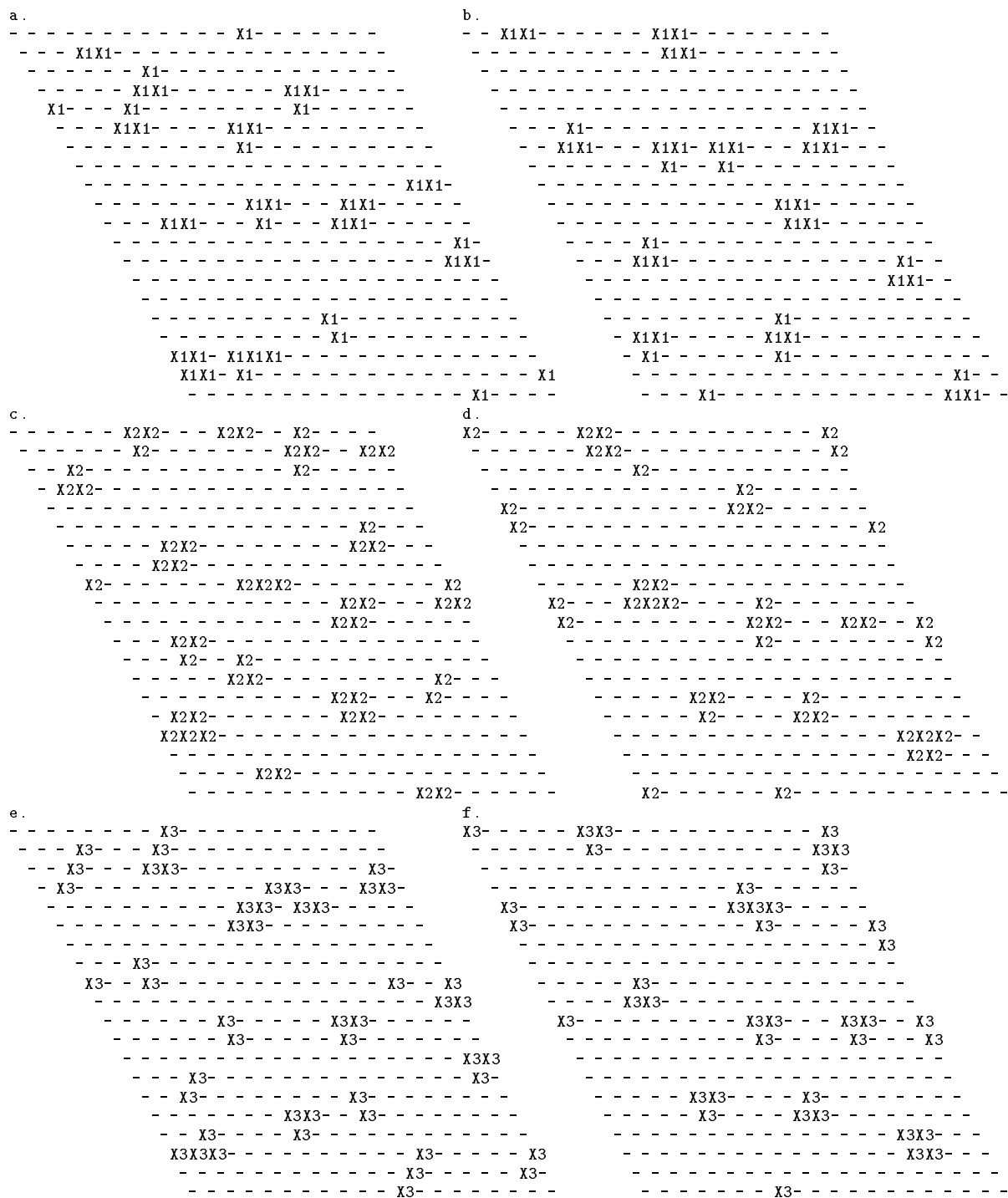


Figure 6.7: The MI visual input maps before (left) and after (right) training. Only the MI input maps for the X dimension are shown (threshold=0.3).

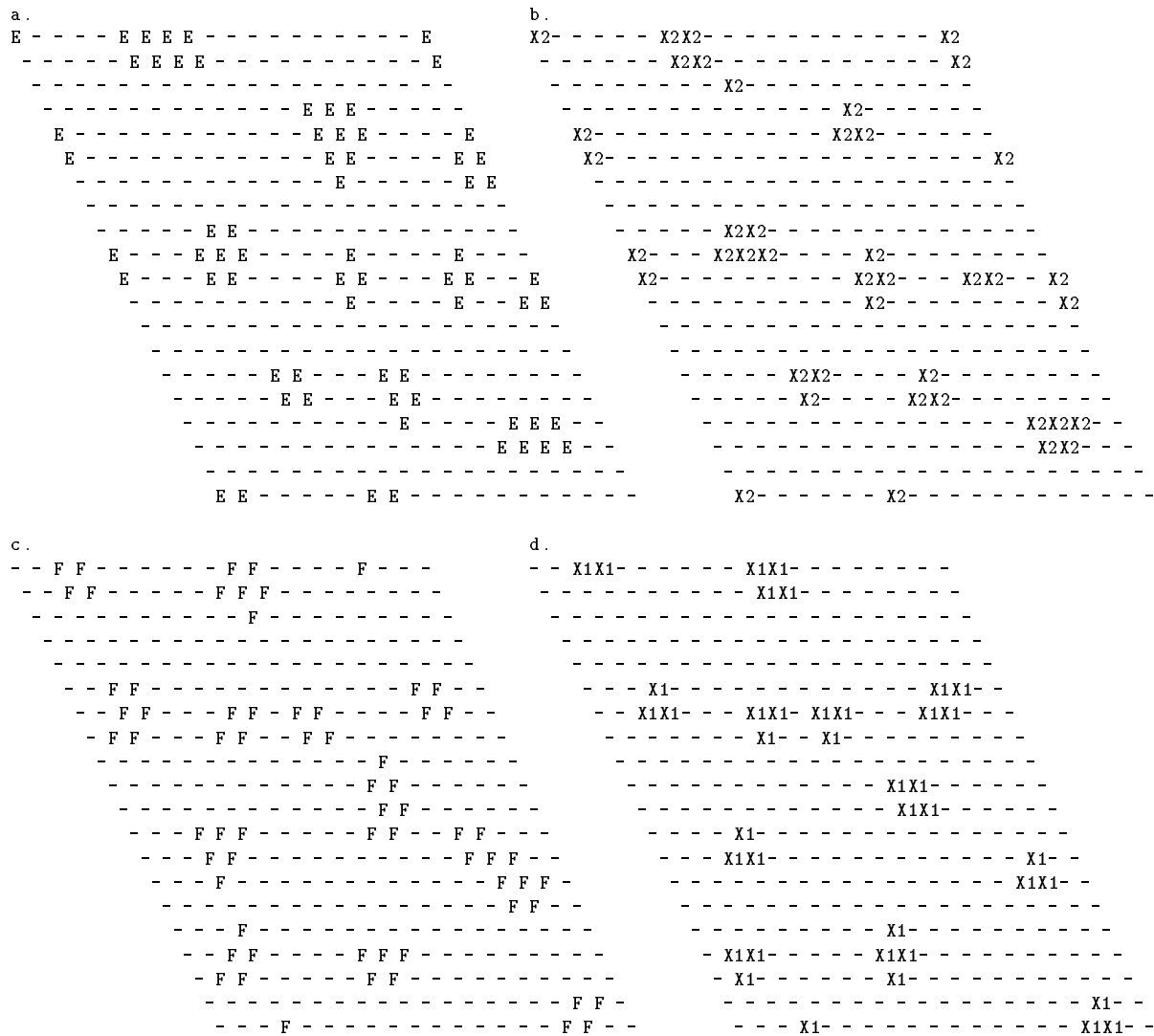


Figure 6.8: Comparison of MI motor output map with MI input map after training. The upper arm extensor (E) of MI output map matches that middle range of the X dimension (X2) of visual input map; The upper arm flexor (F) of MI output map matches the negative range of the X dimension (X1) of visual input maps.

a.

b.

	E	F	B	D	O	C		E	F	B	D	O	C
X1	0.43	0.41	0.41	0.41	0.44	0.43	X1	0.05	<b>0.88</b>	0.11	0.21	0.25	0.14
X2	0.39	0.42	0.41	0.42	0.38	0.42	X2	<b>0.90</b>	0.03	0.05	0.08	0.37	0.15
X3	0.42	0.46	0.43	0.40	0.45	0.42	X3	<b>0.86</b>	0.03	0.05	0.08	0.35	0.09
Y1	0.46	0.43	0.42	0.43	0.43	0.41	Y1	0.04	<b>0.94</b>	0.10	0.09	0.26	0.17
Y2	0.45	0.40	0.42	0.42	0.41	0.41	Y2	0.07	0.05	<b>0.80</b>	0.28	0.06	0.06
Y3	0.41	0.45	0.45	0.43	0.47	0.44	Y3	<b>0.71</b>	0.07	0.06	0.11	<b>0.81</b>	0.06
Z1	0.38	0.40	0.41	0.39	0.41	0.44	Z1	0.11	0.06	0.04	<b>0.91</b>	0.13	0.12
Z2	0.40	0.41	0.41	0.40	0.43	0.40	Z2	0.33	0.17	0.05	0.10	0.21	<b>0.68</b>
Z3	0.41	0.45	0.41	0.43	0.43	0.41	Z3	0.08	0.04	<b>0.95</b>	0.05	0.05	0.04

Table 6.6: Similarity values between visual input and motor output features before (left) and after (right) training. Those values that are bigger than 0.6 are in bold style.

The relationships between visual input maps and the motor output maps in MI were also examined. Due to the nature of the such relationships, the correlations between the visual input features and the motor output features are usually not a one-to-one mapping. Although the visual comparison of maps is viable for some strong correlations, such as Fig. 6.8a and b, or Fig. 6.8c and d, other correlations are not always obviously seen with visual inspection. In such a situation, the similarity measuring method is necessary to make quantitative measurements. Table 6.6 shows the similarity values between visual input and motor output features before and after training. It is clear that before training, motor output maps and visual input maps are randomly correlated, making the similarity values range from 0.38 to 0.46. After training, however, clear correlations formed with greatly diversified similarity values. For intuitive illustration, Fig. 6.9 gives a density plot to represent the similarity values in Table 6.6b. The analysis of these values indicated that after training, certain kinds of correlations between visual input and motor output features have become established, reflecting the likelihood of simultaneous presence of particular pairs of features. For example, when the upper arm abductor muscle (B) is contracted, it will move the arm upward and therefore more likely put the hand in the positive range of the Z dimension (Z3) (please refer to Fig. 5.2 in Page 46 showing axes X, Y and Z relative to a human body). Thus, the similarity value between B and Z3 is very large (0.95) after training, indicating a strong correlation between these two features. In most cases, the correlations in Table 6.6b are similar to those observed in the model with visual input only, as illustrated by Table 5.4. In short, those strongly correlated pairs are (X1, F), (X2, E), (X3, E), (Y1, F), (Y3, E), (Y3, O), (Z1, D) and (Z3, B). In general, when one visual input feature is strongly correlated with the motor output feature of a muscle, it is usually anti-correlated (with similarity values smaller than 0.1) with its antagonist muscle.

However, there are also differences:

1. In the visual input only model, Y2 is simultaneously correlated with B and D, with similarity values of 0.46 and 0.63, respectively (Table 5.4). In this model, Y2 is strongly correlated with B (0.80), but not so strongly correlated with D (0.28).
2. In the previous model, Z2 is weakly correlated with E, F, O and C, with similarity values of

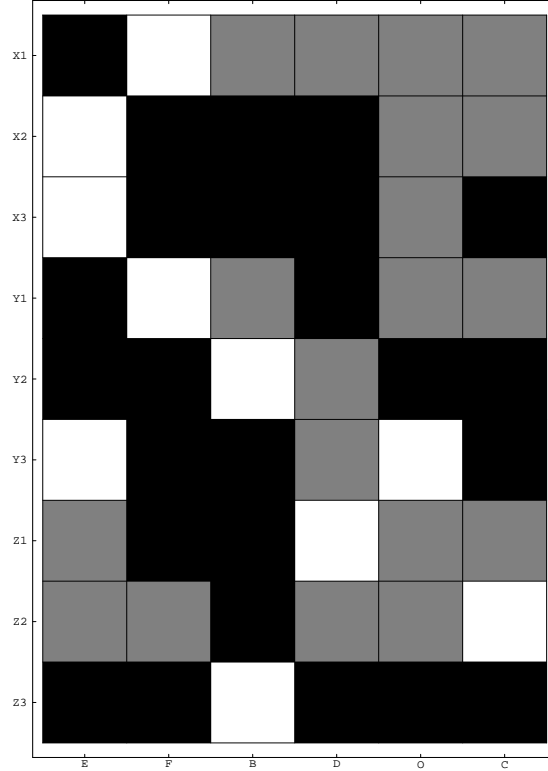


Figure 6.9: A Schematic Density Plot of Table 6.6b. Those similarity values bigger than 0.7 (strongly correlated) are plotted as white blocks in their corresponding positions. The values smaller than 0.1 (anti-correlated) are plotted in black. Others (weak correlations) are plotted in grey.

0.28, 0.23, 0.24 and 0.35, respectively (Table 5.4). In this model, Z2 is strongly correlated with C (0.68). Its similarity values with E, F, O are 0.33, 0.17 and 0.21, respectively.

The above differences indicate that some visual input features that were previously correlated with multiple features now correlate with only one of them. However, this does not mean that every visual input feature is correlated with only one motor output feature in this model. For example, Y3 is correlated with both E and O, with similarity values of 0.71 and 0.81, respectively. On the other hand, the above features (Y2 and Z2) have some spatial symmetric properties. For example, contracting the upper arm abductor (B) or adductor (D) (hence moving the arm up or down) should have the same likelihood of putting the arm in the middle range of Y axis (Y2). There is no reason why Y2 should be correlated stronger with one than the other. It was conjectured that the initial connections and the random training patterns in this individual simulation have caused this biased correlation. Therefore, multiple simulations with different initial connection strengths and training patterns were conducted to confirm this. Table 6.7 shows the similarity values of four additional different simulations of this model, with the same parameters but different initial weights and training patterns. The results are summarized as follows (referring to Fig. 5.2 in Page 46 would be helpful):

1. All of the visual input features other than Y2 and Z2 in these four simulations have the

	E	F	B	D	O	C		E	F	B	D	O	C
X1	0.05	<b>0.92</b>	0.19	0.20	0.22	0.16	X1	0.06	<b>0.91</b>	0.14	0.13	0.10	0.11
X2	<b>0.91</b>	0.03	0.05	0.08	0.44	0.13	X2	<b>0.92</b>	0.07	0.08	0.08	0.78	0.10
X3	<b>0.85</b>	0.02	0.05	0.07	0.37	0.11	X3	<b>0.86</b>	0.07	0.08	0.08	0.72	0.07
Y1	0.05	<b>0.93</b>	0.11	0.14	0.19	0.17	Y1	0.05	<b>0.92</b>	0.07	0.08	0.09	0.14
Y2	0.09	0.13	<b>0.79</b>	0.29	0.07	0.12	Y2	0.08	0.13	0.45	<b>0.70</b>	0.10	0.14
Y3	0.66	0.07	0.05	0.10	<b>0.87</b>	0.11	Y3	<b>0.89</b>	0.08	0.09	0.08	<b>0.89</b>	0.07
Z1	0.09	0.12	0.04	<b>0.92</b>	0.13	0.12	Z1	0.05	0.08	0.07	<b>0.90</b>	0.08	0.10
Z2	0.12	<b>0.78</b>	0.08	0.12	0.38	0.28	Z2	0.31	<b>0.75</b>	0.10	0.08	0.35	0.19
Z3	0.09	0.07	<b>0.94</b>	0.02	0.04	0.09	Z3	0.09	0.14	<b>0.91</b>	0.09	0.09	0.12
	E	F	B	D	O	C		E	F	B	D	O	C
X1	0.03	<b>0.81</b>	0.18	0.25	0.10	0.20	X1	0.05	<b>0.77</b>	0.52	0.14	0.11	0.41
X2	<b>0.93</b>	0.05	0.09	0.06	0.18	0.05	X2	<b>0.79</b>	0.03	0.06	0.16	0.25	0.28
X3	<b>0.89</b>	0.05	0.08	0.06	0.13	0.05	X3	<b>0.84</b>	0.02	0.05	0.16	0.25	0.20
Y1	0.03	<b>0.91</b>	0.12	0.05	0.10	0.20	Y1	0.03	<b>0.87</b>	0.18	0.02	0.09	0.42
Y2	0.04	0.05	0.09	<b>0.94</b>	0.07	0.13	Y2	0.14	0.03	0.16	<b>0.88</b>	0.05	0.21
Y3	0.16	0.09	0.08	0.05	<b>0.94</b>	0.05	Y3	0.39	0.09	0.14	0.07	<b>0.88</b>	0.35
Z1	0.05	0.04	0.04	<b>0.92</b>	0.07	0.08	Z1	0.17	0.03	0.10	<b>0.92</b>	0.06	0.19
Z2	0.12	<b>0.85</b>	0.13	0.05	0.11	0.25	Z2	0.25	0.32	0.08	0.05	0.24	0.46
Z3	0.07	0.18	<b>0.93</b>	0.05	0.10	0.17	Z3	0.05	0.23	<b>0.89</b>	0.08	0.12	0.30

Table 6.7: Similarity values between visual input and motor output features after training in four other simulations with different initial weights and training patterns. Those values that are bigger than 0.7 are in bold style.

same kind of correlations with motor output features as those in Table 6.6. Therefore, it is reasonable to say that these correlations are robust and independent of the initial network condition and training sequences. Also, these correlations are the same as in the model with visual input only, indicating that adding the proprioceptive inputs did not affect these relationships.

- Y2 is strongly correlated with either B or D in different simulations. In all five simulations summarized in Tables 6.6 and 6.7, the similarity values between Y2 and (B,D) are (0.80, 0.28), (0.79, 0.29), (0.45, 0.70), (0.09, 0.94), (0.16, 0.88), respectively. These values indicate that Y2 could be strongly correlated with either B or D, but not both. Recall that in the model with visual input only in Chapter 5, the similarity values between Y2 and (B,D) in five different simulations were (0.46, 0.63), (0.53, 0.47), (0.48, 0.47), (0.45, 0.46), (0.60, 0.46). This means that adding the proprioceptive inputs into the model could prohibit Y2 from being strongly correlated with B and D at the same time. Because of the presence of proprioceptive input, output features of antagonist muscles tend to be anti-correlated, making it impossible for an input feature to be strongly correlated with two anti-correlated feature at the same time.
- Z2 is strongly correlated with one of E, F, O, C. The similarity values between Z2 and (E, F, O, C) in five different simulations are (0.33, 0.17, 0.21, 0.68), (0.12, 0.78, 0.38, 0.28), (0.31, 0.75, 0.35, 0.19), (0.12, 0.83, 0.11, 0.25), (0.23, 0.32, 0.24, 0.46), respectively. In contrast, in the model with visual input only, Y2 is quite evenly correlated with E,F,O,C, with similarity values of five different simulations of (0.23, 0.23, 0.24, 0.35), (0.32, 0.26, 0.24, 0.34), (0.28, 0.31, 0.22, 0.41), (0.28, 0.33, 0.18, 0.34), (0.29, 0.46, 0.31, 0.50), respectively. This means

that adding the proprioceptive inputs into the model could break the balance between evenly correlated features.

In summary, the visual input has no explicit influence on the nature of the proprioceptive input maps and their relationship with motor output maps. On the other hand, the proprioceptive input does influence the visual input maps. With the presence of proprioceptive input, a visual input map can no longer be strongly correlated with a pair of antagonist muscles simultaneously. Even though this visual input has temporal correlation with both of the antagonist pairs, it can only produce spatially correlated maps with one of them.

### 6.2.5 The Relationship between Proprioceptive Input Maps and Visual Input Maps in MI

	E	F	B	D	O	C	e	f	b	d	o	c
X1	<b>0.90</b>	0.00	0.16	0.01	0.03	0.20	0.00	<b>0.90</b>	0.01	0.10	0.37	0.03
X2	0.00	<b>0.85</b>	0.04	0.01	0.06	0.07	<b>0.96</b>	0.00	0.01	0.03	0.07	0.05
X3	0.00	<b>0.78</b>	0.05	0.00	0.01	0.07	<b>0.92</b>	0.00	0.00	0.04	0.09	0.01
Y1	<b>0.97</b>	0.00	0.06	0.00	0.04	0.23	0.00	<b>0.98</b>	0.00	0.01	0.46	0.05
Y2	0.01	0.02	0.20	<b>0.79</b>	0.03	0.14	0.00	0.00	<b>0.80</b>	0.24	0.00	0.00
Y3	0.03	0.38	0.06	0.00	0.01	<b>0.53</b>	<b>0.57</b>	0.03	0.00	0.04	<b>0.53</b>	0.00
Z1	0.01	0.01	<b>0.91</b>	0.02	0.04	0.01	0.02	0.00	0.02	<b>0.96</b>	0.00	0.02
Z2	0.09	0.32	0.12	0.00	<b>0.58</b>	0.07	0.28	0.09	0.00	0.05	0.11	<b>0.59</b>
Z3	0.00	0.02	0.01	<b>0.98</b>	0.02	0.19	0.00	0.00	<b>0.98</b>	0.02	0.00	0.00

Table 6.8: Similarity values between visual input and length (left) and tension (right) proprioceptive input features after training. Those values that are bigger than 0.5 are in bold style.

The relationship between proprioceptive input maps and visual input maps in the MI layer was also studied. The MI layer receives sensory input information from both proprioceptive and visual afferents. The coexistence of proprioceptive input maps and visual input maps in the same cortical layer resulted in interesting relationships between them. In fact, the relationships between proprioceptive and visual input maps can be inferred indirectly from the relationships between proprioceptive input maps and motor output maps and the relationships between visual input maps and motor output maps. It is already known that the MI motor output maps of a particular muscle is aligned with the length input map of its antagonist muscle and the tension map of its own muscle. It is also known that motor output maps form complicated spatial relationships (alignment, partial alignment, mutual exclusion) with visual input maps. As a result, it can be inferred that the correlation of a muscle’s tension input map and a visual input map will be similar to the correlation of the same muscle’s motor output map and the visual input map; while the correlation of a muscle’s length input map and a visual input map will be similar to the correlation of the antagonist muscle’s motor output map and the visual input map. Table 6.8 shows the similarity values between proprioceptive input maps and visual input maps. It is clear that the right half of the table (with a small letter on top of each column to indicate tension) is similar to Table 6.6b. And the left half of the table (with a capital letter on top of each column to indicate length) is similar to Table 6.6b only after each pair of antagonist muscles swap their positions. The detailed

relationships between individual proprioceptive input maps and visual input maps are omitted here, since very similar relationships have been described in previous sections for the visual input maps and motor output maps. Again, these relationships reflect the temporal correlations between input features during the training process. The multiple coexisting sensory input maps and motor output maps in motor cortex form similar relationships so that single layer of cortical elements are able to receive sensory input and send motor output in a consistent way.

### 6.3 Discussion

In the previous sections, it has been shown that the motor control model combining visual and proprioceptive input in motor cortex can form feature maps in the cortical layers during unsupervised learning. By putting activation patterns at randomly selected locations in the motor cortex, the initially random connections self-organized to reflect the arm mechanisms and constraints during training. The self-organized cortical feature maps, along with their relationships, are important in achieving consistent control of arm movement.

This model is different from the previously described motor control models in that it has both proprioceptive and visual inputs. Therefore, it is interesting to see how the combination of these two kinds of input information influence map formation in both proprioceptive and motor cortices. From the results reported earlier, we know that most results are similar to the models with single input pathways alone. In the proprioceptive input pathways, more regularly arranged activation patterns were obtained in the PI and MI layers after training in both the current model and the model with proprioceptive input only. The same kind of relationships were established between length and tension input maps in both models. These results indicate that adding the visual input to the model does not influence the nature of the proprioceptive input maps in a significant way. Moreover, the MI output maps to the lower motor neurons also formed in the same fashion in both models, from initial random maps before training to clusters like formations after training. The correlations between MI proprioceptive input maps and the motor output maps are also the same for both models. This indicates that in the current model, the MI layer, although combining input activations from both proprioceptive and visual pathways, could reconcile the information received and maintain the same correlations between proprioceptive input and motor output maps as the model without visual input. In summary, *adding visual afferent pathways did not significantly affect the self-organization along proprioceptive pathways.*

On the other hand, formation of visual input maps was altered in some ways by the proprioceptive input pathways, when comparing the current motor control model with the model having visual input only. Looking at the visual input maps before and after training did not show much difference between the two models. In both models, the post-training visual input maps were slightly more regular in size of clusters and arrangements. However, the relationships between the visual input maps and the motor output maps were no longer the same for both models, according to the results reported in earlier sections. Particularly, the Y2 feature, which used to be strongly correlated with both B and D at the same time in the model with visual input only, now is only strongly correlated with one of them in each simulation. Whether Y2 is correlated with B or D depends on initial weights and training patterns. From the analysis in Chapter 5, it is presumed that the correlation strength between two features reflects the likelihood of simultaneous presence of these two features during training. Since contracting either the upper arm abductor (B) or the upper arm adductor (D) is going to move the elbow up or down, and is therefore more likely to position the hand in

the middle range of the Y dimension (Y2), it is not surprising to see that Y2 is strongly correlated with B and D. Although B and D are output features of a pair of antagonist muscles, in the model with visual input only, there is no information concerning the relationships of the contraction of antagonist muscles during training. However, after the proprioceptive input pathway was added, certain kind of anti-correlations appeared between mutually antagonist muscles. The lengthening of one muscle will certainly cause shortening of its antagonist muscle. Therefore the length maps of antagonist muscles in both PI and MI layers became mutually exclusive (i.e. no overlapping at all) after training. Due to the relationships between the MI proprioceptive input maps and motor output maps, the motor output maps of mutually antagonist muscles also became mutually exclusive. This means that an element in the MI layer is not likely to activate both B and D at the same time. On the other hand, training will causes Y2 to be associated with B and D, because of the frequent occurrence of features pairs (Y2,B) or (Y2,D) during training. The network tried to compromise this contradiction by letting Y2 be strongly correlated with either B or D, depending on the initial condition of the network. Similar things also happened to the correlations between Z2 and E,F,O,C. In the model with visual input only, Z2 was weakly correlated with one of the four muscles, showing no strong correlations with any individual one. In the model with combined proprioceptive and visual inputs, the the correlations of Z2 with any pair of antagonist muscles at the same time were discouraged, causing one of the four muscle (E,F,O,C) to be strongly correlated with Z2. In summary, *adding proprioceptive inputs changed the visual input maps' correlation with the motor output maps, eliminating the situations where one visual input feature correlated simultaneously with two output features of antagonist muscles.* The influence of adding proprioceptive inputs is to differentiate the antagonist muscles, causing anti-correlated input and output maps for antagonist muscles. This model tells us that although visual input alone can produce consistent input-output maps, it is not able to identify antagonist muscle pairs; proprioceptive input is necessary to help distinguish the antagonist muscles.



## Chapter 7

# Some Analysis of Cortical Feature Map Formation

In previous chapters, a number of versions of motor control models were presented. Cortical feature map formation in these models and map relationships were examined. It has been shown that cortical feature maps form spontaneously in the models via unsupervised learning, and the relationships in these maps reflected the internal characteristics of the model's control task. Hence it would be interesting to explore the mathematical basis of map formation when multiple maps are present simultaneously. For example, one could ask the question of why the observed map relationships form via training, and under what conditions such relationships form. This chapter tries to answer this type of question from a simplified mathematical point of view.

Because of the complexity of the motor control model, and the non-linear dynamics of the mathematical formulas imposed on the model, it may appear virtually impossible to do a mathematical analysis in a comprehensive fashion. Nevertheless, analysis of some aspects of the model, as well as analysis of a simplified version of the model, can still give us insight into the driving force behind the emergent phenomena.

This chapter has primarily four parts. First, an abstract model will be described in order to simplify the analysis. Second, the nature of output activation patterns with respect to single input activation will be studied. It can be shown that under certain conditions, the activations of output units converge to an equilibrium point over time under competitive distribution of activation. Also, the activation of an output unit is a monotonically increasing function of its relative connection strength from the original input unit. In the third and fourth parts, the study focuses on how the temporal correlations between input features can be transformed into spatial correlations of the output feature maps during training. It is found that when two input units are in perfect temporal correlation, their output feature maps are in perfect spatial correlation after training. On the other hand, when two input units are in perfect anti-correlation, it may produce anti-correlated spatial feature maps only under some restrictions.

## 7.1 The Abstract Model

The analysis in this chapter will be based on an abstract model, so that the analysis is more mathematically tractable and the results are widely applicable. This abstract model has only two layers of units: one layer of input units, and one layer of output units (Fig. 7.1). The input units and output units are fully connected, unless otherwise specified. The input units have no lateral connections. The output units may or may not have lateral connections, depending on the situations discussed in later sections.

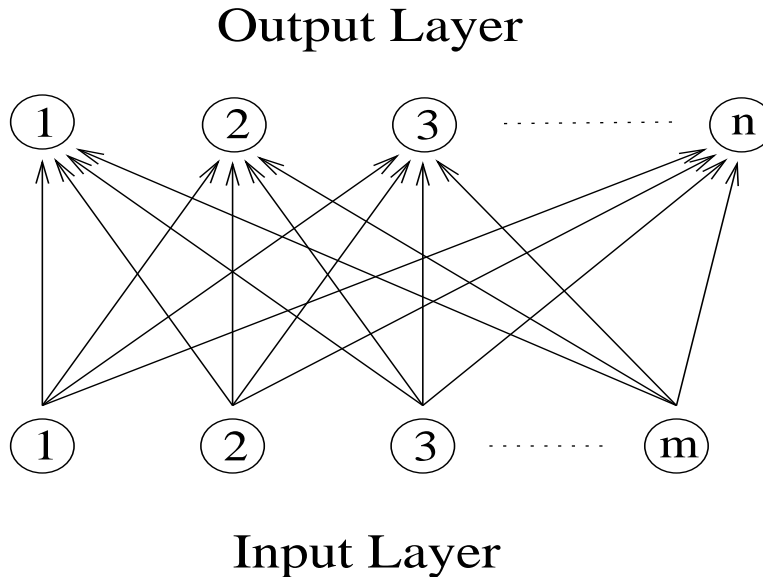


Figure 7.1: The abstract model for mathematical analysis of cortical feature map relationships. There are  $m$  units in the input layer, which are fully connected with the  $n$  units in the output layer.

There are  $m$  input units and  $n$  output units. The activation level of input units are designated as  $I_1, I_2, \dots, I_m$ . The activation levels of output units are designated as  $a_1, a_2, \dots, a_n$ . The connection strength from input unit  $i$  to output unit  $j$  is  $w_{ji}$ , for  $i = 1, 2, \dots, m$  and  $j = 1, 2, \dots, n$ . The units in the input layer receive initial activations:  $\mathbf{I} = (I_1, I_2, \dots, I_m)$ . An input activation pattern is held steady while activation spreads to the output units, using competitive distribution of activation as designated in Equation 3.8 and 3.9. After the activation in output units stabilizes, a competitive learning rule (Equation 3.10 and 3.11) is used to update the weights  $w_{ji}$  for  $i = 1, 2, \dots, m$  and  $j = 1, 2, \dots, n$ .

In this model, each input unit is considered to be an input feature. A certain feature is present if and only if the corresponding input unit is activated. Multiple features can be present simultaneously, as they were in most cases in the training input patterns for the motor control model. There may be some temporal relationship between different input features in the training patterns. The measurement of an output map with respect to a particular input feature is done by activating that input unit and measuring the activation pattern in the output units. The analysis in this chapter will focus on the relationship of the output maps for different input features. We will study how the temporal correlations between input features are translated into spatial correlations in the output feature maps.

## 7.2 Activation Patterns of Output Units

In this section, the analysis is focused on the activation patterns in the output units of the abstract model described earlier, using competitive distribution of activation. Competitive distribution of activation is a more dynamic activation rule than standard activation rules. There is no

general theory about the convergence of activation in different models using competitive distribution of activation. In this analysis, it is assumed that initially all input and output units have zero activation, and then an input pattern  $\mathbf{I}$  is applied to the input layer and held steady. The activation spreads from input units to output units according to the rule of competitive distribution of activation, and finally stabilizes. The study that follows will focus on the convergence of the activation patterns and the relationship between the final activation pattern in the output layer and the input activation pattern. It will be shown that in certain simplified situations, activation in the output layer converges to equilibrium; the resultant activation levels can be derived as a closed form formula; and the final activation of a particular output unit is a monotonic function of incoming connection strength. These results not only provide us with a better understanding about the nature of the activation rule under competitive distribution of activation, but also provide theoretical background for directly using weights to measure cortical maps under certain conditions, as described earlier in this dissertation. The results in this section are also used in the analysis in later sections.

There are several past analyses of self-organized feature maps. Most of these are based on Kohonen's model. Since Kohonen's model uses a simple rule to compute activation of output units, the analyses are focused on the convergence of training, not the convergence of activation. In the one dimensional case, it has been proved that the Kohonen's algorithm converges [Kohonen, 1989; Kohonen, 1995; Erwin *et al.*, 1992; Lo *et al.*, 1993]. For higher dimensions, the results are only partial [Ritter & Schulten, 1986; Erwin *et al.*, 1992].

There is also some previous work about the final activation patterns using competitive distribution of activation. [Benaim & Samuelides, 1990; Reggia & Edwards, 1990] used different versions of activation rules that imposed no bounds on activation levels, and focused their analysis on the conditions for getting stable activation levels. [Wang & Seidman, 1988] used an activation rule that has an upper bound of 1.0 which can be an equilibrium point. That study was focused on equilibrium of activation in several different models, based on different initial activation patterns. In these analyses there were no external inputs. Also no activation of any unit was held steady. The network relied on its own initial activation and activation rule to reach a stable position. The activation rule used in the analysis in this section (as well as in our previous simulations with the motor control model) is different from the above activation rules in that it has an upper bound for the activation level, and this upper bound can never become an equilibrium point. Also some portion of the network (namely input units) are held steady and corresponding activations in the output layer are determined. In summary, past related analyses all used a different activation rule than that used here.

Given a pattern  $\mathbf{I} = (I_1, I_2, \dots, I_m)$  (assuming  $I_i > 0$ ) clamped on the inputs, a weight matrix  $\mathbf{W} = \{w_{ji}\}$ , and initial activation values for output units, the activation levels of output units  $\mathbf{a}(\mathbf{t}) = (a_1(t), a_2(t), \dots, a_n(t))$  are uniquely determined at any given time, according to the competitive distribution of activation rule. Since the measurement of a cortical map of any particular feature is determined by stimulating single input layer units and measuring the corresponding activations in the output layer, the analysis in this section is limited to the situation where only one input unit is activated. That is:  $\mathbf{I} = (I_1, I_2, \dots, I_m)$  where  $I_j \neq 0$  for a particular unit  $j$  and all other  $I_i = 0$  for  $i \neq j$ . The problem can be further simplified by assuming the absence of lateral connections in the output layer. The following analysis is based on this assumption.

Now consider output unit  $k$ . According to the activation rule and the assumption of having only one input unit activated, we have:

$$\frac{da_k(t)}{dt} = c_s a_k(t) + (M - a_k(t)) in_k(t) \quad (7.1)$$

where

$$in_k(t) = out_{kj}(t) = c_p \frac{(a_k(t) + q)w_{kj}}{\sum_l (a_l(t) + q)w_{lj}} I_j. \quad (7.2)$$

Here  $a_k(t)$  is the activation level of output unit  $k$ , and  $in_k(t)$  is the activation received by unit  $k$  from the input layer, which equals  $out_{kj}(t)$ , the activation from input unit  $j$  to output unit  $k$ , because unit  $j$  is the only activated unit in the input layer. Parameter  $c_s < 0$  is the decay constant indicating how fast activation decays, and  $M > 0$  is the maximum value of activation. Parameter  $c_p > 0$  is the output gain constant, determining the fraction of activation to be output. The parameter  $q$  has two effects: one is to prevent division by zero when initial activations in output units are zero;  $q$  can also be used to control the degree of competition. Since in the simulations of the motor control model demonstrated in previous chapters a very, very small  $q$  was used most of the time and it did not significantly contribute to the control of competitiveness, we will assume that  $q$  values are zero in this analysis and that there are very small, equal initial activations in the output units to avoid division by zero. Hence Equation 7.2 becomes (omitting time  $t$  for convenience):

$$in_k = c_p \frac{a_k w_{kj}}{\sum_l a_l w_{lj}} I_j \quad (7.3)$$

Substituting Equation 7.3 into Equation 7.1, we have:

$$\frac{da_k}{dt} = c_s a_k + (M - a_k) c_p \frac{a_k w_{kj}}{\sum_l a_l w_{lj}} I_j \quad (7.4)$$

Equation 7.4 is actually a set of differential equations, for  $k = 1, 2, \dots, n$ , with initial small but equal positive values for every  $a_k(0)$ . Before we derive the activation levels at equilibrium, we must first prove that this set of differential equations will reach equilibrium when time  $t$  goes to infinity, instead of oscillating forever, or some other behavior. To prove the convergence of Equation 7.4 is a little complicated. Note that Equation 7.4 can be rewritten as:

$$\frac{da_k}{dt} = (c_s + (M - a_k) c_p \frac{w_{kj}}{\sum_l a_l w_{lj}} I_j) a_k, \quad (7.5)$$

emphasizing that  $a_k = 0$  is a possible equilibrium point. To eliminate the effect when  $a_k$  is close to zero, we first examine an altered differential equation set:

$$\frac{da_k}{dt} = c_s + (M - a_k) c_p \frac{w_{kj}}{\sum_l a_l w_{lj}} I_j \quad (7.6)$$

We will show that Equation 7.5 converges to a set of values that satisfy:

$$c_s + (M - a_k) c_p \frac{w_{kj}}{\sum_l a_l w_{lj}} I_j = 0 \quad (7.7)$$

for all  $k = 1, 2, \dots, n$ . Any solution of Equations 7.7 is an equilibrium point of Equations 7.6. What we need to prove is the system will converge to this solution from the beginning point. Note that Equations 7.6 will not converge on every beginning point. For example, if  $a_k < 0$  for all  $k = 1, 2, \dots, n$ , then Equations 7.6 simply diverge.

In the following, we assume that the same conditions hold as held in the computational studies described throughout this chapter. Specifically, we assume that the decay constant  $c_s < 0$ , the excitatory gain constant  $c_p > 0$ , the maximum activation  $M > 0$ , and that external inputs are  $I_j > 0$ , i.e., excitatory. We also assume that initially,  $w_{ij} > 0$  and  $0 < a_i(0) < M$  for all  $i$  and  $j$ . These conditions were always satisfied by the models described in this dissertation.

**Lemma 7.1** *If the above conditions are satisfied, then*

(a)  $a_i(t) < M$  for any  $t \geq 0$

(b)  $\sum_{l=1}^n a_l w_{lj} > 0$  for any  $t \geq 0$

will hold on the trajectory governed by Equations 7.6.

Proof: Part (a) is apparent from Equations 7.6 with the assumption that  $a_i(0) < M$ . For part (b), it is apparent that  $\sum_{l=1}^n a_l w_{lj} > 0$  at  $t = 0$  by the assumption in the text above. Activations  $a_k$  move on the trajectory decided by Equations 7.6. When  $\sum_{l=1}^n a_l w_{lj}$  becomes sufficiently close to zero, its reciprocal will become very big. As a result, the right hand side of Equations 7.6 will become positive, necessarily causing  $a_k$  to increase. Therefore  $\sum_{l=1}^n a_l w_{lj}$  will increase and move away from zero.  $\square$

To prove the convergence of Equations 7.6, we first prove a two-dimensional special case. Then we generalize the result to the n-dimensional situation.

**Proposition 7.1** *The set of differential equations in Eq. 7.6 converge to a set of values that satisfy Equation 7.7 in a two dimensional space (i.e.  $n = 2$ ).*

Proof: We need to prove the convergence of:

$$\begin{cases} \frac{da_1}{dt} = c_s + (M - a_1)c_p \frac{w_{1j}}{a_1 w_{1j} + a_2 w_{2j}} I_j \\ \frac{da_2}{dt} = c_s + (M - a_2)c_p \frac{w_{2j}}{a_1 w_{1j} + a_2 w_{2j}} I_j \end{cases} \quad (7.8)$$

Let:

$$\begin{aligned} L1 &= c_s + (M - a_1)c_p \frac{w_{1j}}{a_1 w_{1j} + a_2 w_{2j}} I_j \\ L2 &= c_s + (M - a_2)c_p \frac{w_{2j}}{a_1 w_{1j} + a_2 w_{2j}} I_j \end{aligned}$$

with some algebra,  $L1 = 0, L2 = 0$  can be rewritten as:

$$\begin{cases} w_{1j}(1+A)a_1 + w_{2j}a_2 = w_{1j}AM \\ w_{1j}a_1 + w_{2j}(1+A)a_2 = w_{2j}AM \end{cases} \quad (7.9)$$

Here  $A = \frac{c_p}{-c_s} I_j$  is a positive constant. Calculate the slope of  $L1 = 0, L2 = 0$  in the space  $(a_1, a_2)$ :

$$\begin{cases} Slope(L1 = 0) = -\frac{w_{1j}(A+1)}{w_{2j}} \\ Slope(L2 = 0) = -\frac{w_{1j}}{w_{2j}(A+1)} \end{cases} \quad (7.10)$$

Since  $w_{1j}, w_{2j}$  and  $A$  are all positive, we know that both  $Slope(L1 = 0)$  and  $Slope(L2 = 0)$  are negative. Also since  $A > 0$ ,  $|Slope(L1 = 0)| > |Slope(L2 = 0)|$ . As a result,  $L1 = 0$  is always steeper than  $L2 = 0$  (see Fig. 7.2). Also  $|Slope(L1 = 0)| > |Slope(L2 = 0)|$  implies that  $L1$  and  $L2$

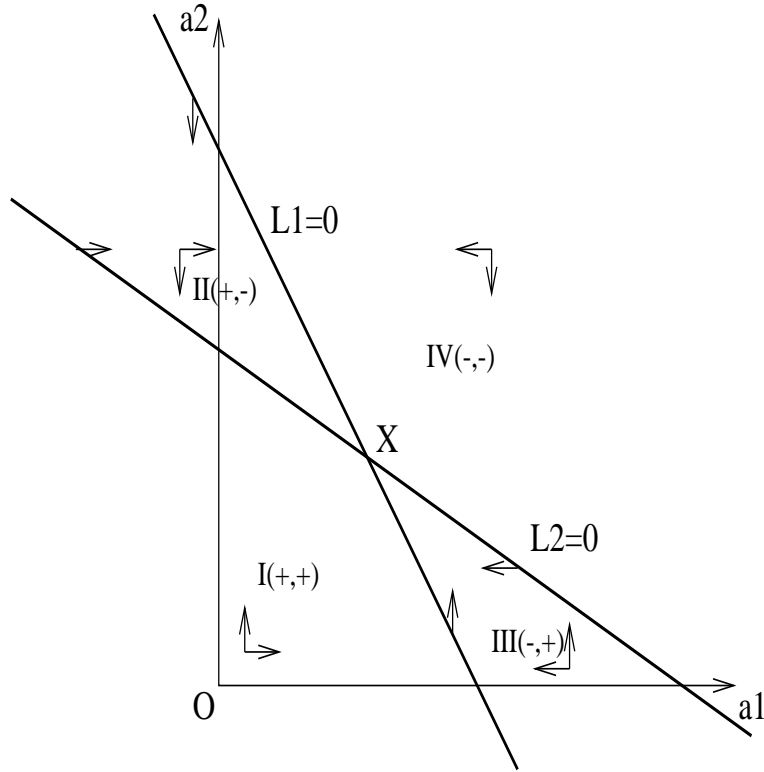


Figure 7.2: Drawing of  $L1 = 0$  and  $L2 = 0$  in the two dimensional space  $(a_1, a_2)$ .  $L1 = 0$  and  $L2 = 0$  intersect at  $X$ .  $L1 = 0$  and  $L2 = 0$  divide the plane into four sections: I, II, III and IV. Each section has a unique combination of the sign of  $L1$  and  $L2$ , indicated as  $+$  and  $-$ . The small arrows indicate the moving direction of either  $a_1$  or  $a_2$  in each section as well as on the boundaries.

are not parallel. So  $L1$  and  $L2$  have a unique intersection point. Draw  $L1 = 0, L2 = 0$  in the plane of  $(a_1, a_2)$ , it will be similar to Fig. 7.2. The intersection of  $L1 = 0$  and  $L2 = 0$ ,  $X$ , may not be in the first quadrant. It may be in the second or fourth quadrant.

$L1 = 0$  and  $L2 = 0$  divide the plane into four sections, named as I, II, III and IV. We only consider the area where  $a_1 w_{1j} + a_2 w_{2j}$  is greater than zero, which includes entire first quadrant and part of the second and fourth quadrant. Each of these four sections has a unique combination of the sign of  $L1$  and  $L2$ , indicated as  $+$  and  $-$  in Fig. 7.2. Since  $\frac{da_1}{dt} = L1$  and  $\frac{da_2}{dt} = L2$ , the signs of  $L1$  and  $L2$  also determine the direction that the system  $(a_1, a_2)$  moves when governed by Equation 7.8. The small arrows in Fig. 7.2 indicate the moving direction of either  $a_1$  or  $a_2$  in each section as well as on the boundaries. It is not difficult to see that, when  $(a_1, a_2)$  is in any of the four sections, the sign of  $\frac{da_1}{dt}$  and  $\frac{da_2}{dt}$  will make sure that it moves towards the intersection point  $X$ . For example, in section II,  $L1 > 0$  and  $L2 < 0$ , so  $a_1$  is increased and  $a_2$  is decreased, moving  $(a_1, a_2)$  towards  $X$ . Also on the lines  $L1 = 0$  and  $L2 = 0$ , the movement directions are either horizontal or vertical, as indicated by small arrows on the lines (see Fig. 7.2). As a result,  $(a_1, a_2)$  will finally converge to  $X$  no matter where it starts.

Note that the drawing of Fig. 7.2 depends on the slope of  $L1 = 0$  and  $L2 = 0$ . It is important to have  $L1 = 0$  be steeper than  $L2 = 0$ . Otherwise the signs in Fig. 7.2 will be different and  $(a_1, a_2)$

could move away from  $X$ , and the system would not converge.  $\square$

We have proved that Equation 7.8 will converge to the equilibrium point of  $X$  in two dimensional space. Now we extend the result into  $n$ -dimensional space.

**Proposition 7.2** *The set of differential equations in Equation 7.6 converge to a set of values that satisfy Equation 7.7.*

Proof: First, arbitrarily pick a dimension  $i$ . For a given point  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  in  $n$ -dimensional space, define:

$$L_i(\mathbf{a}) = c_s + (M - a_i)c_p \frac{w_{ij}}{\sum_l a_l w_{lj}} I_j \quad (7.11)$$

We will show that  $\mathbf{a}$  will move on the trajectory governed by Equation 7.6 towards the point where  $L_i(\mathbf{a}) = 0$ .

First, assume that  $L_i(\mathbf{a}) > 0$ . In this case,  $a_i$  will increase over time. We will show that  $\mathbf{a}$ 's trajectory will decrease  $L_i$  to zero. There are two situations: i) for all other  $k \neq i$ ,  $\frac{da_k}{dt} > 0$ ; ii) for some  $k$ ,  $\frac{da_k}{dt} \leq 0$ .

*Case 1:* For all other  $k \neq i$ ,  $\frac{da_k}{dt} > 0$ . That means every component of  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  is increasing. Then by Equation 7.11,  $L_i(\mathbf{a})$  is decreasing over time. This trend continues until either  $L_i(\mathbf{a}) = 0$  or one of  $a_k$  no longer increases. The latter occurrence reduces the situation to the next case.

*Case 2:* For some  $k$ ,  $L_k = \frac{da_k}{dt} \leq 0$ . Fix the values of  $a_l$  for  $l \neq i, k$ . We study the subspace defined by  $a_i$  and  $a_k$  (see Fig. 7.3).  $L_i = 0$  and  $L_k = 0$  can be rewritten as:

$$\begin{cases} w_{ij}(1+A)a_i + w_{kj}a_k = w_{ij}AM - \sum_{l \neq i,k} a_l w_{lj} \\ w_{ij}a_i + w_{kj}(1+A)a_k = w_{kj}AM - \sum_{l \neq i,k} a_l w_{lj} \end{cases} \quad (7.12)$$

The slopes of  $L_i = 0$  and  $L_k = 0$  are:

$$\begin{cases} Slope(L_i = 0) = -\frac{w_{ij}(A+1)}{w_{kj}} \\ Slope(L_k = 0) = -\frac{w_{kj}}{w_{ij}(A+1)} \end{cases} \quad (7.13)$$

We have  $|Slope(L_i = 0)| > |Slope(L_k = 0)|$ . Draw the projection of hyperplane of  $L_i = 0$  and  $L_k = 0$  in subspace  $(a_i, a_k)$  as indicated in Fig. 7.3. Since  $L_i > 0$  and  $L_k \leq 0$ , the projection of  $\mathbf{a}$  in  $(a_i, a_k)$  subspace should be in section II. In this section, the increase of  $a_i$  tends to move  $\mathbf{a}$  closer to  $L_i = 0$ , but  $a_k$  decreases to move  $\mathbf{a}$  downward. Whether  $\mathbf{a}$  will move closer to  $L_i = 0$  is decided by relative value of  $L_i$  and  $L_k$ . However, as indicated in Fig. 7.3, the boundary between section II and I has only positive  $L_i$  and  $L_k$  is zero. As a result,  $\mathbf{a}$  can only be pushed towards the corner  $X$ , and therefore moves towards  $L_i = 0$ . It should be noted that  $\mathbf{a}$  not only moves in  $(a_i, a_k)$  subspace, but also moves in other dimension at the same time. Fig. 7.3 is true in  $(a_i, a_k)$  subspace for any  $a_l$  ( $l \neq i, k$ ). That is, in every  $(a_i, a_k)$  subspace,  $\mathbf{a}$  is going to be pushed into a corner similar to  $X$ . Moreover, for other dimensions  $l \neq i, k$ , there is a similar cross plane drawing. Either  $L_l > 0$  and  $\mathbf{a}$  move towards  $L_l = 0$ , or  $L_l \leq 0$  and  $\mathbf{a}$  is pushed into the corner of section II. As a result,  $\mathbf{a}$  will moves to  $L_i = 0$  no matter which dimension it moves.

The argument of  $L_i < 0$  is very similar to the above and  $\mathbf{a}$  will move towards  $L_i = 0$ . Since  $i$  is an arbitrary dimension,  $\mathbf{a}$  will eventually moves to the intersection where  $L_i = 0$  for all  $i = 1, 2, \dots, n$ .

$\square$

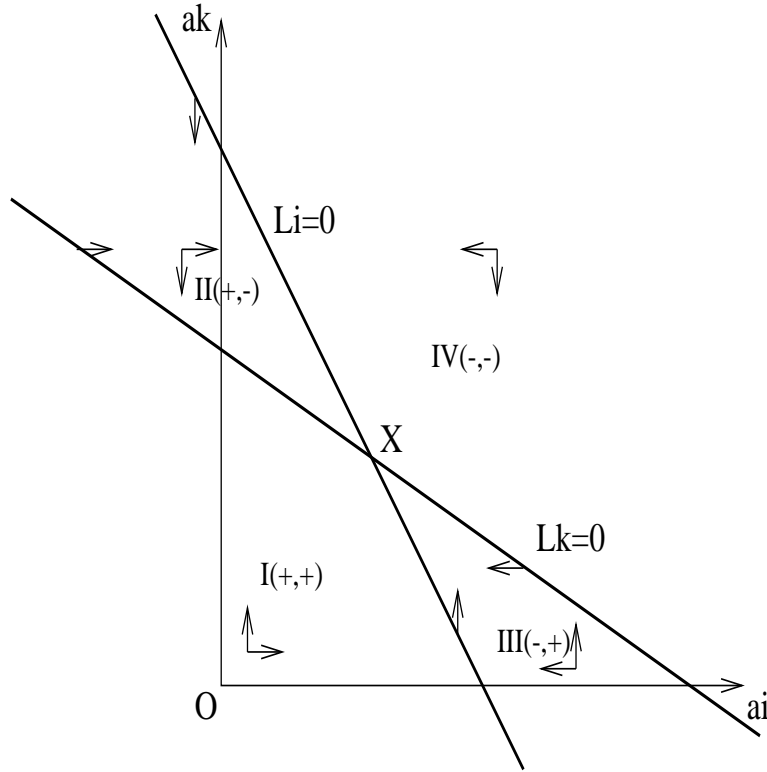


Figure 7.3: Drawing of  $L_i = 0$  and  $L_k = 0$  in subspace  $(a_i, a_k)$  of the  $n$ -dimensional space. The projection of hyperplane  $L_i = 0$  and  $L_k = 0$  in  $(a_i, a_k)$  subspace intersect at  $X$ .  $L_i = 0$  and  $L_k = 0$  divide the subspace into four sections: I, II, III and IV. Each section has a unique combination of the sign of  $L_i$  and  $L_k$ , indicated as  $+$  and  $-$ . The small arrows indicate the moving direction of either  $a_i$  or  $a_k$  in each section as well as on the boundaries.

So far we have proved the convergence of differential set in Eq. 7.5. Now we come back to prove the original differential equation set in Eq 7.4.

**Proposition 7.3** *The set of differential equations in Eq. 7.4 converge to an equilibrium point.*

Proof: Consider the intersection point  $X$  of hyper planes  $L_i = 0$  for  $i = 1, 2, \dots, n$ .  $X$  has coordinates  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  that satisfies Equations 7.7. There are two different situations:

*Case 1:*  $X$  is in the first quadrant. That means  $x_k > 0$  for all  $k = 1, 2, \dots, n$ .

Comparing Equations 7.5 with Equations 7.6, it is apparent that the only difference is that Equations 7.5 has one more factor:  $a_k$  on the right hand side of the equation. Since initially every  $a_k$  has a small positive value, adding such a factor will not change the sign on the right hand side of the equation. So the sign Equation 7.5 will be always the same as that of Equation 7.6 for all  $k = 1, 2, \dots, n$ . Based on Proposition 7.2, the point  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  will move towards the intersection point  $X$ . Since  $X$  is in the first quadrant, the sign of  $a_k$  is never going to change during the movement towards  $X$ . Finally, the point  $\mathbf{a}$  will reach the equilibrium point  $X$ , where all  $\frac{da_k}{dt}$  become zero.

*Case 2:*  $X$  is not in the first quadrant. That means there are some  $x_k < 0$ .



Again, with initial small positive values for all  $a_k$ , the point  $\mathbf{a}$  is in the first quadrant and will move towards intersection point  $X$ , just like the previous case. However, since  $X$  is not in the first quadrant,  $\mathbf{a}$  will move across the quadrant boundary at some point. When this happens, the factor  $a_k$  in Equation 7.5 will take into effect. For example, suppose for a particular  $m$ ,  $x_m < 0$ , and point  $\mathbf{a}$  has move to a place where  $a_m$  become zero. In this case, the equation:

$$\frac{da_m}{dt} = (c_s + (M - a_m)c_p \frac{w_{mj}}{\sum_l a_l w_{lj}} I_j) a_m \quad (7.14)$$

will be dominated by the factor  $a_m$ , making  $\frac{da_m}{dt}$  be zero and preventing  $a_m$  from becoming negative. As a result, Equation 7.14 has reached its equilibrium point and  $a_m$  will no longer change. From the neural network point of view, unit  $m$  has zero activation and therefore has quit the competition from the competitive distribution of activation. From a spatial geometry point of view, the system degenerates from  $n$ -dimension to  $(n-1)$ -dimension. The point will continue to move in the hyperplane of  $a_m = 0$ , and all the hyperplanes  $L_k$  also project into this hyperplane, forming an  $(n-1)$ -dimensional system. This degenerate process will continue until all dimensions that have a negative  $x_k$  have  $a_k = 0$ . In the remaining subspace, the intersection point  $X$  will be in the first quadrant and the subsystem will converge to that point.  $\square$

Now that we have proved the convergence of the Equations 7.4. It is not too difficult to find out the stabilizing values of this equation set.

**Proposition 7.4** *The stabilizing values of the differential equations in Eq. 7.4 are:*

$$a_k = M \left( 1 - \frac{\sum_l w_{lj}}{(A + n)w_{kj}} \right) \quad (7.15)$$

for  $k = 1, 2, \dots, n$ , provided that this set of values correspond to the point in the first quadrant of the  $n$ -dimensional space.

Proof: We already know that Equations 7.4 will converge. When the system reaches the equilibrium point, we have:

$$c_s a_k + (M - a_k) c_p \frac{a_k w_{kj}}{\sum_l a_l w_{lj}} I_j = 0 \quad (7.16)$$

Here, we can safely assume that  $a_k \neq 0$ . So dividing Eq. 7.16 by  $a_k$ , we have:

$$c_s + (M - a_k) c_p \frac{w_{kj}}{\sum_l a_l w_{lj}} I_j = 0 \quad (7.17)$$

Rearranging Eq. 7.17, we have:

$$(A + 1)w_{kj}a_k + \sum_{l \neq k} a_l w_{lj} = M A w_{kj} \quad (7.18)$$

Here  $A = \frac{c_p}{-c_s} I_j$ . Eq. 7.18 holds for every  $k = 1, 2, \dots, n$ . So this is basically a set of linear equations:

$$\begin{cases} (A + 1)w_{1j}a_1 + w_{2j}a_2 + \dots + w_{nj}a_n = M A w_{1j} \\ w_{1j}a_1 + (A + 1)w_{2j}a_2 + \dots + w_{nj}a_n = M A w_{2j} \\ \vdots \\ w_{1j}a_1 + w_{2j}a_2 + \dots + (A + 1)w_{nj}a_n = M A w_{nj} \end{cases} \quad (7.19)$$

Solving this equation set, we have:

$$a_k = M\left(1 - \frac{\sum_l w_{lj}}{(A+n)w_{kj}}\right) \quad (7.20)$$

for  $k = 1, 2, \dots, n$ .  $\square$

This result indicates that the activation level in output unit  $k$  with respect to stimulation at a particular input unit  $j$  is a monotonically increasing function of the “relative connection strength” between unit  $j$  and  $k$ . Here “relative connection strength” between unit  $j$  and  $k$  is the ratio of connection strength between unit  $j$  and  $k$  to the summation of the connection strengths between unit  $j$  to all the output units. The larger proportion  $w_{kj}$  has among  $w_{lj}$  (for  $l = 1, 2, \dots, n$ ), the higher the activation  $a_k$  will be. Also the parameter  $A = (c_p * I_j)/(-c_s)$  will affect activation level, the more output gain ( $c_p$ ) or less decay ( $c_s$ ), the higher  $a_k$  would be.

Equations 7.15 are only valid when the intersection point is in the first quadrant. In case that for some  $k$ ,  $a_k = M\left(1 - \frac{\sum_l w_{lj}}{(A+n)w_{kj}}\right) < 0$ , the system will degenerate, according to Proposition 7.3. In such case  $w_{kj}$  is too weak to keep any positive  $a_k$ . As a result,  $a_k$  remains zero and quit the competition. In this situation, we have following proposition.

**Proposition 7.5** *Without loss of generality, assume that weights are sorted so that  $w_{1j} > w_{2j} > \dots > w_{nj}$ . Also assume that  $h$  is the largest number in  $1, 2, \dots, n$  such that equations*

$$c_s + (M - a_k)c_p \frac{w_{kj}}{\sum_{l=1}^h a_l w_{lj}} I_j = 0 \quad (7.21)$$

for  $k = 1, 2, \dots, h$  have an intersection in the first quadrant. The equilibrium values of the differential equations in Eq. 7.4 are

$$a_k = M\left(1 - \frac{\sum_l w_{lj}}{(A+h)w_{kj}}\right) \quad (7.22)$$

for  $k = 1, 2, \dots, h$ , and  $a_k = 0$  for  $k = h + 1, \dots, n$ .

Proof: Equations 7.20 tell us that the unit with the smallest weight has the smallest activation. With  $w_{kj}$  already sorted in descending order, and all the initial  $a_k$  (for  $k = 1, 2, \dots, n$ ) are the same small positive number, it is apparent that  $(c_s + (M - a_n)c_p \frac{w_{nj}}{\sum_{l=1}^n a_l w_{lj}} I_j) a_n$  is the smallest (the most negative) among  $1, 2, \dots, n$ . As a result,  $a_n$  is the first activation level to approach zero. Variable  $a_n$  then remains at zero, reducing the system to  $(n-1)$ -dimensions. This process continues until we find a largest number  $h$  such that Equations 7.21 have all positive solutions. At this point, the system no longer degenerates. By following the same derivation procedure, we have:

$$a_k = M\left(1 - \frac{\sum_l w_{lj}}{(A+h)w_{kj}}\right) \quad (7.23)$$

for  $k = 1, 2, \dots, h$ .  $\square$

Here a simple example is described to illustrate the numerical results of above analyses. Assume there is one input node and three output nodes in the model. The connection strengths are  $w_1 = 0.3, w_2 = 0.2, w_3 = 0.1$ , respectively. Also assume the input node has activation of 1.0 and is held steady. The decaying constant  $c_s$  is  $-1.0$ ; the output gain constant  $c_p$  is 1.0. We first calculate the intersection point of  $L_1 = 0, L_2 = 0$ , and  $L_3 = 0$ . By using Equation 7.15, the intersection

point can be calculate as  $a_1 = 0.5, a_2 = 0.25, a_3 = -0.5$ . It turns out that this point is not in the first quadrant of  $(a_1, a_2, a_3)$  space . It is negative in  $a_3$  dimension. In real simulation,  $a_1, a_2$  and  $a_3$  have initial positive values, and  $\mathbf{a} = (a_1, a_2, a_3)$  will move toward the intersection point. When  $a_3$  reaches zero, it no longer change, according to Equation 7.5. The system degenerates into a two dimensional system. The intersection of  $L_1 = 0, L_2 = 0$  is  $a_1 = 0.444, a_2 = 0.167$ . This point is in the first quadrant of  $(a_1, a_2)$  space. So the system converge to this point. As a result, the final activation of the output nodes in equilibrium point is:  $a_1 = 0.444, a_2 = 0.167, a_3 = 0$ . The numerical calculation with difference equations of this sample model yields the same result.

So far, we have obtained the result about activation levels of output units at equilibrium, with respect to single input stimuli. This result can be used in further analysis of the formation of cortical feature maps, since the measurement of cortical feature maps is conducted by stimulating only one input unit at a time, and measuring the corresponding activation pattern in the output units. Some of the analysis later in this chapter will use this result.

The result obtained here also helps us to have a better understanding of the relationship between the output activation patterns and the corresponding connection strengths. With a single activated input unit, some output units with weak connections from this stimulated input unit will have zero activation, while other output units compete for activation based on their relative connection strength with this input unit. From Equation 7.22, it is apparent that the activation level of an output unit is a monotonically increasing function of the relative connection strength of this unit. This observation provides theoretical background for using weight vectors directly as the measurement of a cortical feature map. Note that in previous chapters, when motor output maps of MI were measured, instead of stimulating MI units and measuring the activation patterns in lower motor units, the weight vectors were used directly as a substitute for such activation patterns. From the results obtained in this section, it is clear that the order of the magnitude will be the same regardless of using weight vectors or activation patterns. Therefore, such a simplification of map measurements would not affect the appearance of the resultant maps, as long as appropriate thresholds were selected.

It should be noted that above simplified measurement method was only used in measuring the output map from MI to lower motor neurons, where no lateral connections exist between units in the output layer. It will not apply to other map measurement situations such as cortical input maps, where output units (or more precisely, measured units) were laterally connected. In the latter situation, the network dynamics become more complicated and the activation of a particular output unit not only depends on its connection strength from input units, but also depends on the activation of adjacent units in the same layer.

### 7.3 The Relationship of Cortical Maps with Perfectly Correlated Features

In this section and the next section, the analysis will be focused on the relationship between different feature in cortical feature maps. The simulation results presented in the motor control models in previous chapters have indicated that the formed cortical feature maps have features that exhibit certain kind of relationship between each other. In general, it was found that via training, *temporally* correlated input features usually form *spatially* correlated feature maps in the same cortical layer. Such kind of correlation is subject to a theoretical investigation.

Because of the complexity and dynamics of this model, a comprehensive theoretical analysis would be extremely difficult, if possible. For simplicity, two special cases are studied. One case is to study the relationships between cortical feature maps with two input features that are in perfect correlation (correlation coefficient 1.0). The other case is to study the relationships between cortical feature maps with two input features that are in perfect anti-correlation (correlation coefficient -1.0). These are two extreme cases; all other cases fall somewhere in between. The relationships between the length proprioceptive input maps and the tension proprioceptive input maps in both PI and MI described in previous chapters usually reflect these two special cases. While the relationships between visual input maps and proprioceptive input maps sometimes reflect intermediate situations. The relationships between sensory input maps and motor output maps in the original motor control model can be regarded as a special form of relationships between input maps, as the output signals are fed back as sensory signals through the closed-loop system. Therefore the analysis results of the two extreme cases in the simplified model can give us significant insight into the correlations of cortical feature maps in general.

In this section, the analysis is focused on the case where two input features are in perfect correlation. Suppose input units  $i$  and  $j$  are in perfect correlation. That is:  $I_i = I_j$  for any input pattern  $\mathbf{I}^1$ . There is no restriction or assumption about the values of other input units. We study the correlation of features  $i$  and  $j$ . We will show that perfect temporal correlation in input features can produce perfect spatial correlated feature maps in output layer.

**Proposition 7.6** *Suppose that for every input pattern  $\mathbf{I}$ ,  $I_i = I_j$ , for an arbitrary but particular  $i$  and  $j$ . Also suppose that during training, each output unit has an unlimited number of chances of being activated. Then after training, the cortical feature maps of  $i$  and  $j$  become identical (or fully aligned).*

Proof: Consider the weight changes from unit  $i$  and  $j$  to a particular output unit  $k$  during the training process. According to the competitive learning rule, we have:

$$w_{ki}^{new} = w_{ki}^{old} + \eta[I_i - w_{ki}^{old}]a_k^* \quad (7.24)$$

$$w_{kj}^{new} = w_{kj}^{old} + \eta[I_j - w_{kj}^{old}]a_k^* \quad (7.25)$$

where  $\eta$  is the learning rate constant. The quantities  $w_{ki}^{old}$ ,  $w_{kj}^{old}$  and  $w_{ki}^{new}$ ,  $w_{kj}^{new}$  are weights from unit  $i$ ,  $j$  to unit  $k$  before and after a learning cycle, respectively.  $a_k^*$  is the thresholded activation defined as:

$$a_k^* = \begin{cases} a_k - \alpha & \text{if } a_k > \alpha \\ 0 & \text{otherwise} \end{cases} \quad (7.26)$$

where  $\alpha$  is a threshold constant. We want to show that during training, the  $w_{ki}$  and  $w_{kj}$  get closer and closer in value, and finally become equal. To calculate the difference between  $w_{ki}$  and  $w_{kj}$ , subtract Equation 7.25 from Equation 7.24:

$$w_{ki}^{new} - w_{kj}^{new} = w_{ki}^{old} - w_{kj}^{old} + \eta[I_i - I_j - (w_{ki}^{old} - w_{kj}^{old})]a_k^* \quad (7.27)$$

Since  $I_i = I_j$ , we have:

---

<sup>1</sup>Actually, when the correlation coefficient between  $I_i$  and  $I_j$  is 1.0, it only means a perfect linear dependency between  $I_i$  and  $I_j$ . In case that both  $I_i$  and  $I_j$  are in range  $[0, 1]$ , and assuming  $I_i$  and  $I_j$  have same expectation (i.e. both have same chance to get activated), we can safely say that  $I_i = I_j$ .

$$|w_{ki}^{new} - w_{kj}^{new}| = |1 - \eta a_k^*| |w_{ki}^{old} - w_{kj}^{old}| \quad (7.28)$$

Here using absolute values insures that  $w_{ki}$  and  $w_{kj}$  will get closer regardless of which one is bigger. In order to have  $|w_{ki}^{new} - w_{kj}^{new}| < |w_{ki}^{old} - w_{kj}^{old}|$ , we must have  $|1 - \eta a_k^*| < 1$ . We know that  $\eta$  is a positive constant usually much smaller than 1.  $a_k^*$  is a non-negative number smaller than maximum activation constant  $M$ . So a sufficient small learning rate  $\eta$  will ensure that  $1 - \eta a_k^* > 0$ , and hence  $|1 - \eta a_k^*| \leq 1$ . In case that  $a_k^* = 0$ , we have  $|w_{ki}^{new} - w_{kj}^{new}| = |w_{ki}^{old} - w_{kj}^{old}|$ . We have assumed that during training, there are an infinite number of chances for  $a_k^*$  to be greater than zero<sup>2</sup>, then  $|w_{ki}^{new} - w_{kj}^{new}| < |w_{ki}^{old} - w_{kj}^{old}|$  holds, and after sufficient training patterns, we have  $|w_{ki}^{new} - w_{kj}^{new}|$  close to 0. This is true for every output unit  $k$ . As a result, we end up having identical weight vectors  $\mathbf{w}_{*i} = (w_{1i}, w_{2i}, \dots, w_{ni})$  and  $\mathbf{w}_{*j} = (w_{1j}, w_{2j}, \dots, w_{nj})$ . Since the cortical feature maps of features  $i$  and  $j$  only depend on the incoming weight vector from input unit  $i$  and  $j$ , respectively, identical weight vectors will lead to identical (or fully aligned) cortical feature maps.  $\square$

In the above, it has been shown that the temporally correlated input features will make their corresponding weight vectors close to each other, and therefore lead to spatially fully aligned cortical feature maps. However, in an actual simulation of the motor control model, the modification of weights is a little bit more complicated. Not only the competitive learning rule is applied, but also a normalization procedure is applied after each learning cycle, to avoid weight vectors growing without limit. We ignored this issue above. The following proposition will take this normalization procedure into account.

**Proposition 7.7** *Perfectly correlated input in input units  $i$  and  $j$  will lead to fully aligned cortical feature maps of  $i$  and  $j$  even when a normalization process is used during learning, provided that each output unit has an unlimited number of chances to be activated.*

Proof: For any output unit  $k$ , the sum of all of the components of the incoming weight vector is kept constant all the time during the training process. That is:

$$\sum_l w_{kl} = S \quad (7.29)$$

where  $S$  is a given constant. After the competitive learning rule is applied to the weights, a renormalization procedure occurs to insure Equation 7.29 still holds. Assume that  $w_{kl}^{old}$  is the weight from unit  $l$  to unit  $k$  at the beginning of a given learning cycle,  $w_{kl}^{new}$  is the weight after competitive learning rule is applied, and  $w_{kl}^{new'}$  is the weight after renormalization. We have:

$$w_{kl}^{new} = w_{kl}^{old} + \eta [I_l - w_{kl}^{old}] a_k^* \quad (7.30)$$

and

$$w_{kl}^{new'} = \frac{S}{\sum_m w_{km}^{new}} w_{kl}^{new}$$

---

<sup>2</sup>More precisely, we should say there are infinite number of chances such that  $M > a_k^* \geq \epsilon > 0$ , where  $\epsilon$  is any given small positive constant.

$$\begin{aligned}
&= \frac{w_{kl}^{old} + \eta(I_l - w_{kl}^{old})a_k^*}{\sum_m [w_{km}^{old} + \eta(I_m - w_{km}^{old})a_k^*]} S \\
&= \frac{w_{kl}^{old} + \eta(I_l - w_{kl}^{old})a_k^*}{[S + \eta(\sum_m I_m - S)a_k^*]} S
\end{aligned}$$

Therefore

$$\begin{aligned}
|w_{ki}^{new'} - w_{kj}^{new'}| &= \left| \frac{w_{ki}^{old} + \eta(I_i - w_{ki}^{old})a_k^*}{[S + \eta(\sum_m I_m - S)a_k^*]} S - \frac{w_{kj}^{old} + \eta(I_j - w_{kj}^{old})a_k^*}{[S + \eta(\sum_m I_m - S)a_k^*]} S \right| \\
&= \left| \frac{S(1 - \eta a_k^*)}{[S + \eta(\sum_m I_m - S)a_k^*]} \right| |w_{ki}^{old} - w_{kj}^{old}| \\
&= \left| \frac{1 - \eta a_k^*}{1 + \frac{\sum_m I_m}{S} \eta a_k^* - \eta a_k^*} \right| |w_{ki}^{old} - w_{kj}^{old}|
\end{aligned}$$

after rearranging the formula, we have:

$$|w_{ki}^{new'} - w_{kj}^{new'}| = \left| \frac{1}{1 + \frac{\eta a_k^* \sum_m I_m}{1 - \eta a_k^* S}} \right| |w_{ki}^{old} - w_{kj}^{old}| \quad (7.31)$$

From Equation 7.31, it is clear that  $\frac{\eta a_k^* \sum_m I_m}{1 - \eta a_k^* S}$  must be greater than zero in order for

$$|w_{ki}^{new} - w_{kj}^{new}| < |w_{ki}^{old} - w_{kj}^{old}| \quad (7.32)$$

to hold. Here  $S$  is a positive constant. And  $\sum_m I_m$  is presumably greater than zero.  $1 - \eta a_k^*$  is also greater than zero when  $\eta$  is small and  $a_k^*$  is in normal range. So Equation 7.32 will hold if and only if  $\eta a_k^*$  is greater than zero.  $\square$

Thus, a similar conclusion could be reached as the case without weight normalization: when each output unit has an infinite number of chances to be activated above threshold during the training process, and sufficient training patterns are provided, perfect temporal correlation of input features will lead to fully aligned cortical feature maps. It should be pointed out that the results obtained from the analysis above are based on relatively few restrictions. The above analysis does not rely on specific activation rule, as long as that activation rule can produce positive activation with certain upper bound. Therefore the result of analysis should apply in systems with different activation rules. In fact, this widely applicable result is due to the nature of the competitive learning rule. The competitive learning rule modifies the weights in a way such that activated output units have their incoming weight vectors shift towards the input vector, in the  $m$  dimensional input space. After training the incoming weight vector of a particular output unit points to the average position of the set of input vectors that cause it to activate. This property of the competitive learning rule not only holds in the entire input space, but also holds in any of its subspaces. That is, in any subspace, the incoming weight vector (in that subspace) of a unit will still shift towards the input pattern (in that subspace). Hence we can hand pick dimension  $i$  and  $j$  to form a two dimensional subspace. In this subspace, a particular output unit will have its input weight vector pointing to

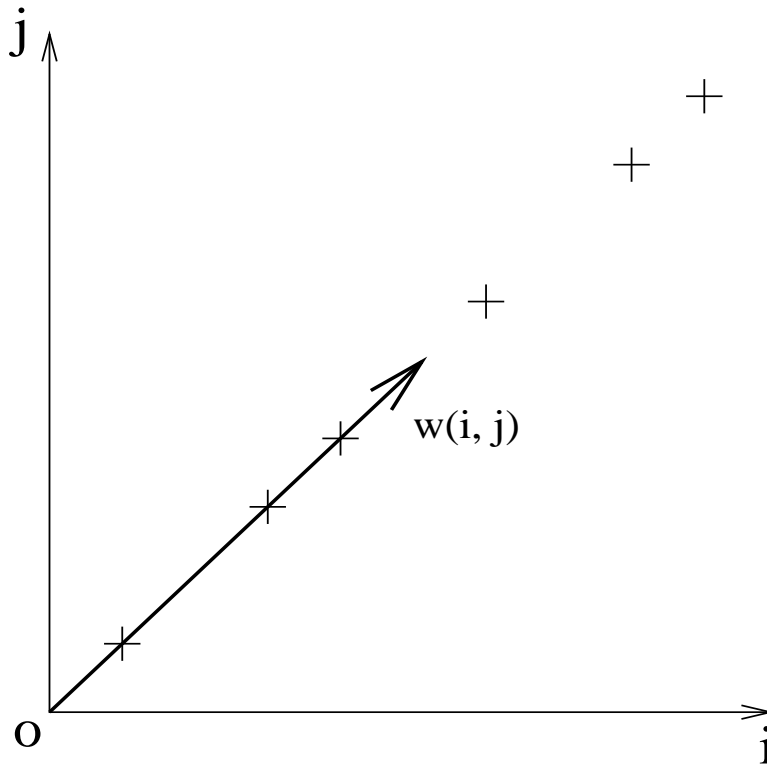


Figure 7.4: A subspace (in dimension  $i$  and  $j$ ) of the  $m$  dimensional input space. The crosses in the figure shows the positions of input patterns in this subspace that can activate a particular output unit. Since all these input patterns have their  $i$ 'th and  $j$ 'th components equal, the final incoming weight vector  $w(i, j)$ , which points to the average position of those input patterns, will also lie along the diagonal line.

the average position of the input vectors that activated it during training. This average position will lie on the diagonal line of the two dimensional subspace, as indicated in Fig. 7.4, because all of the input patterns (and hence all the input patterns that could activate this output unit) have their  $i$ 'th and  $j$ 'th components equal. Note that different output units may have different lengths of incoming weight vectors in this subspace. Some may be long enough to meet the threshold during map measurement, others may not. Yet they all lie on the diagonal line. Therefore, after every output unit has been activated a sufficient number of times, the output maps with respect to input features  $i$  and  $j$  will become fully aligned to each other.

## 7.4 The Relationship of Cortical Maps with Perfectly Anti-correlated Features

In the previous section, it has been shown that perfectly correlated input features will generate identical output maps after training. In this section, we look at the other extreme case: two input features that are in perfect anti-correlation. That is: the correlation coefficient between two inputs values at unit  $i$  and  $j$  is  $-1$ . In a simple case, the input unit  $i$  and  $j$  has one and only one activated

in each training pattern, namely  $\langle a_i, a_j \rangle \in \{\langle 1, 0 \rangle, \langle 0, 1 \rangle\}$ . Based on simulation results reported in previous chapters, one would anticipate a theoretical conclusion that anti-correlated input features will produce corresponding feature maps that are mutually exclusive. That is: there is no output unit tuned to both features. This turns out not to be true. In order for an output unit  $k$  to be tuned to either stimulation in input unit  $i$  or  $j$ , but not both, it is necessary to have  $w_{ki}$  and  $w_{kj}$  to segregate during training. However, depending on the parameter setting and initial weights, it is possible that  $w_{ki}$  and  $w_{kj}$  do not segregate during training, but converge to become the same value. As a result, unit  $k$  can become tuned to both unit  $i$  and  $j$ . When all of the output units have equal weights from unit  $i$  and  $j$ , the maps for feature  $i$  and  $j$  become identical, instead of mutually exclusive. In this section, we will focus our analysis on the condition under which the output units segregate their incoming weights from unit  $i$  and  $j$ , hence producing mutually exclusive feature maps.

In a simple case, we study a network with no input units other than units  $i$  and  $j$  (or we can assume other input units exist but always have input values of zero) in order to avoid interference from other input units. The network uses competitive distribution of activation along with a competitive learning rule, and is repetitively presented with input patterns of  $\langle 1, 0 \rangle$  or  $\langle 0, 1 \rangle$  during the training. In this case, we have the following proposition:

**Proposition 7.8** *A sufficient condition for the network to produce mutually exclusive feature maps after training is that the parameter setting satisfies:*

$$A = \frac{c_p}{-c_s} < n \quad (7.33)$$

where  $n$  is the number of output units,  $c_p$  is the output gain, and  $c_s$  is decay constant.

Proof: First consider the changes of the weights during the training. We start with the cost function (or Lyapunov function) associated with the competitive learning rule. The cost function has the form:

$$E\{w_{ki}\} = \frac{1}{2} \sum_{\mu} a_k^{\mu*} (I_i - w_{ki})^2 \quad (7.34)$$

where  $a_k^{\mu*}$  is the thresholded activation as indicated by Equation 7.26, and  $\mu$  is an input pattern. Here  $\mu \in \{\langle 1, 0 \rangle, \langle 0, 1 \rangle\}$ .

The gradient of the cost function is given by:

$$-\eta \frac{\partial E}{\partial w_{ki}} = \eta \sum_{\mu} a_k^{\mu*} (I_i - w_{ki}) \quad (7.35)$$

Since  $\mu \in \{\langle 1, 0 \rangle, \langle 0, 1 \rangle\}$ , Equation 7.35 becomes:

$$-\eta \frac{\partial E}{\partial w_{ki}} = \eta a_k^{\langle 1, 0 \rangle*} (1 - w_{ki}) + \eta a_k^{\langle 0, 1 \rangle*} (0 - w_{ki}) \quad (7.36)$$

The gradient of the cost function points to the direction in the weight space that would minimize the cost function. So, we define:

$$\Delta w_{ki} = \eta a_k^{\langle 1, 0 \rangle*} (1 - w_{ki}) + \eta a_k^{\langle 0, 1 \rangle*} (0 - w_{ki}) \quad (7.37)$$



The following analysis will use this equation instead of the conventional individual update rule, because Equation 7.37 reflects the real direction that the weights shift towards. This equation also corresponds to so-called batch mode competitive learning rule ([Hertz *et al.*, 1993]), in which case each of the input patterns is presented to the network and the changes of the weights are accumulated and then updated altogether.

Similar to Equation 7.37, for  $w_{kj}$ , we have:

$$\Delta w_{kj} = \eta a_k^{<1,0>*} (0 - w_{kj}) + \eta a_k^{<0,1>*} (1 - w_{kj}) \quad (7.38)$$

Also it is assumed that initially  $w_{ki}$  and  $w_{kj}$  are not equal, otherwise this batch mode learning will always yield the same value of  $w_{ki}$  and  $w_{kj}$ , after each iteration of learning<sup>3</sup>. With this assumption, it is also safe to assume that  $w_{ki} > w_{kj}$  for the purpose of simplifying formulas in the following analysis (assuming otherwise will not affect analysis results). We need to study the condition under which after each iteration of learning,  $w_{ki}$  and  $w_{kj}$  will segregate further. With all the weights to be positive and  $w_{ki} > w_{kj}$ , we want to have:

$$w_{ki}^{new} - w_{kj}^{new} > w_{ki}^{old} - w_{kj}^{old} \quad (7.39)$$

So from Equation 7.39, we have

$$\begin{aligned} \Delta w_{ki} - \Delta w_{kj} &= (w_{ki}^{new} - w_{ki}^{old}) - (w_{kj}^{new} - w_{kj}^{old}) \\ &= (w_{ki}^{new} - w_{kj}^{new}) - (w_{ki}^{old} - w_{kj}^{old}) \\ &> 0 \end{aligned} \quad (7.40)$$

It should also be noted that  $w_{ki} + w_{kj} = 1$  will hold during the training process if this incoming weight vector was initialized to be 1 and each input pattern also has same length. Therefore

$$\begin{aligned} \Delta w_{ki} - \Delta w_{kj} &= \eta a_k^{<1,0>*} (1 - w_{ki}) + \eta a_k^{<0,1>*} (0 - w_{ki}) \\ &= \eta a_k^{<1,0>*} (0 - w_{kj}) + \eta a_k^{<0,1>*} (1 - w_{kj}) \\ &= 2\eta (w_{kj} a_k^{<1,0>*} - w_{ki} a_k^{<0,1>*}) \end{aligned} \quad (7.41)$$

Here we are going to use the analysis results in Proposition 7.4. Assuming all the output units are nonzero, we have

$$a_k^{<1,0>*} = M \left( 1 - \frac{\sum_l w_{li}}{(A+n)w_{ki}} \right) \quad (7.42)$$

$$a_k^{<0,1>*} = M \left( 1 - \frac{\sum_l w_{lj}}{(A+n)w_{kj}} \right) \quad (7.43)$$

where  $A = \frac{c_p}{-c_s}$  as  $I_i^{<1,0>} = I_j^{<0,1>} = 1$ .

Plug Equation 7.42 and 7.43 into Equation 7.41, we have

$$\Delta w_{ki} - \Delta w_{kj} = 2M\eta \left[ \left( \frac{w_{ki}}{w_{kj}} \frac{\sum_l w_{lj}}{(A+n)} - \frac{w_{kj}}{w_{ki}} \frac{\sum_l w_{li}}{(A+n)} \right) - (w_{ki} - w_{kj}) \right] \quad (7.44)$$

---

<sup>3</sup> Actually, an individual update rule can break this equality. So it is safe to assume  $w_{ki} \neq w_{kj}$ ; otherwise simply apply individual update rule once before using batch mode learning.

Note that  $\sum_l w_{li} + \sum_l w_{lj} = \sum_l (w_{li} + w_{lj}) = n$ . When weights are initially random and  $n$  is relatively large,  $\sum_l w_{li}$  and  $\sum_l w_{lj}$  have similar values. Hence  $\sum_l w_{li} \approx \sum_l w_{lj} \approx n/2$ . As a result, Equation 7.44 becomes

$$\Delta w_{ki} - \Delta w_{kj} \approx 2M\eta \left[ \left( \frac{w_{ki}}{w_{kj}} - \frac{w_{kj}}{w_{ki}} \right) \frac{n}{2(A+n)} - (w_{ki} - w_{kj}) \right] \quad (7.45)$$

Since we want  $\Delta w_{ki} - \Delta w_{kj} > 0$ , we only need to have

$$\left( \frac{w_{ki}}{w_{kj}} - \frac{w_{kj}}{w_{ki}} \right) \frac{n}{2(A+n)} - (w_{ki} - w_{kj}) > 0 \quad (7.46)$$

Equation 7.46 will ensure that after one iteration of  $\{< 1, 0 >, < 0, 1 >\}$  learning,  $w_{ki}$  and  $w_{kj}$  will segregate further. However, it will not guarantee that in the next iteration of learning,  $w_{ki}$  and  $w_{kj}$  will also segregate even further. To ensure that, we need

$$\left( \frac{w_{ki}}{w_{kj}} - \frac{w_{kj}}{w_{ki}} \right) \frac{n}{2(A+n)} - (w_{ki} - w_{kj}) \quad (7.47)$$

to be monotonically increasing with respect to  $w_{ki}$ , so that when  $w_{ki}$  becomes larger (namely segregates further away from  $w_{kj}$ ) in one iteration, it will become even larger in the next iteration of learning. To simplify representation, let's define  $w_{ki} = x$ , so  $w_{kj} = 1 - x$ . Then Equation 7.47 becomes a function

$$f(x) = \left( \frac{x}{1-x} - \frac{1-x}{x} \right) \frac{n}{2(A+n)} - (x - (1-x)) \quad (7.48)$$

where  $x \in (0.5, 1]$ . For  $f(x)$  to be monotonically increasing,  $f'(x) > 0$  must hold. So

$$f'(x) = \left( \frac{1}{(1-x)^2} + \frac{1}{x^2} \right) \frac{n}{2(A+n)} - 2 > 0 \quad (7.49)$$

which is equivalent to

$$\frac{1}{(1-x)^2} + \frac{1}{x^2} > \frac{4(A+n)}{n} \quad (7.50)$$

Since

$$\frac{1}{(1-x)^2} + \frac{1}{x^2} > \frac{1}{0.5^2} + \frac{1}{0.5^2} = 8 \quad (7.51)$$

for  $x \in (0.5, 1]$ . So we let

$$8 > \frac{4(A+n)}{n} \quad (7.52)$$

After simplification, we have

$$A < n \quad (7.53)$$

This equation will guarantee that  $f(x)$  will be monotonically increasing. Also it is not difficult to prove that when Equation 7.53 holds, Equation 7.46 will also hold. Therefore, whenever there is a slight difference between the initial  $w_{ki}$  and  $w_{kj}$ , and unit  $k$  can be activated, the difference

will be come bigger and bigger during training. Eventually the unit will tune to only one of the  $\{< 1, 0 >, < 0, 1 >\}$  patterns. When all the output units can be activated, the cortical maps with respect to input feature  $i$  and  $j$  will be mutually exclusive.  $\square$

Plug  $A = \frac{c_p}{-c_s}$  into Equation 7.53, we have

$$\frac{c_p}{-c_s} < n \quad (7.54)$$

This equation illustrates the conditions that the parameters of the network must satisfy in order to produce mutually exclusive feature maps: (1) use smaller output gain  $c_p$ ; (2) use bigger magnitude of decaying constant  $|c_s|$ ; or (3) use more output units. These conditions give us some insight into network behavior besides empirical experiments.

There are several considerations here. First, it should be noted that  $n$  here actually represent the number of units that can be activated. Depending on the weight setting, some output units get such weak input activation that they may never get activated and therefore have zero activation. The larger  $n$  is in the above condition (3), the more output units that have sufficiently strong connections so that they can participate in the competition. Second, the above derivation of condition assumes that a certain output unit  $k$  is activated for both  $< 1, 0 >$  and  $< 0, 1 >$  patterns. This is true only when both  $w_{ki}$  and  $w_{kj}$  are relatively strong, namely close to 0.5. On the other hand, we are actually only concerned about this situation. That is, whether a pair of initially close connections to the same output unit ( $w_{ki}$  and  $w_{kj}$ ) will segregate through training and eventually becomes tuned to only one of the input patterns. For units that have an initial segregated connections from inputs  $i$  and  $j$  (i.e., one of them is close to 1, the other is close to 0), the above analysis is not valid as in each iteration of  $\{< 1, 0 >, < 0, 1 >\}$  learning, the units are not activated all the time. However, this is also less of a concern, because these units have already formed mutually exclusive maps. Third, from the derivation process, it is clear that Equation 7.54 is a sufficient condition, not a necessary condition. It is possible that, when Equation 7.54 does not hold, the weights still segregate through training. This is most related to the initial values of  $w_{ki}$  and  $w_{kj}$ . Equation 7.54 only ensures that any  $w_{ki}$  and  $w_{kj}$  will segregate through training, unless they are both exactly 0.5. Fourth, the above analysis uses gradient descent rule that similar to the learning in a batch mode fashion, where each one of the two input patterns is presented to the network alternatively. In fact, the analysis results also apply to the usual incremental learning rule, where input patterns are presented to the network in random order, as long as the input patterns have an equal probability of being used. In such a case the gradient descent will actually point to the same direction. One advantage of using incremental learning rule is that it can force segregation of weights even when they are initially all 0.5. Fifth, Equation 7.53 indicates that for a given network, the combination of  $\frac{c_p}{-c_s}$  is a single indicator of convergence. Individual values of  $c_p$  and  $c_s$  are not important, as long as this ratio is kept within certain range.

Fig. 7.5 shows a very simple network as an example to illustrate the conditions discussed above that will ensure the segregation of incoming weights. In this example, the number of output units  $n = 2$ . According to Equation 7.53,  $A < 2$  will ensure the segregation of weights. If we use Equation 7.46 and plug in the corresponding weight values, we have  $A < \frac{13}{6} \approx 2.17$ . To make sure that function  $f(x)$  in Equation 7.48 is monotonically increasing, where  $x = w_{ki} = 0.6$  here, we have:  $A < \frac{181}{72} \approx 2.51$ . So, combining the two conditions, we get a more precise condition:  $A = \frac{c_p}{-c_s} < 2.17$ . This condition is slightly less restrictive than Equation 7.53, but it depends on this particular set of weights.

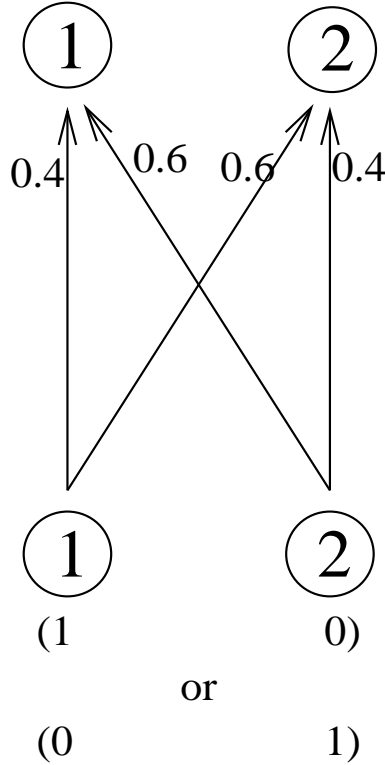


Figure 7.5: An example of a small network with only two input units and two output units. The numbers on the connections indicate the weights. When presenting alternate input patterns of  $(1,0)$  and  $(0,1)$  to the network, the two output units may tune to different input patterns, depending on the parameter setting.

It should also be noted that the analysis results obtained in this section are based on a variety of assumptions. First, the competitive distribution of activation is used as the activation rule, since this is used in our motor control system. Other activation rules may yield different results. And these results can be derived starting from Equation 7.41 and plugging in corresponding activation rule. Second, it is assumed that there are no lateral connections in output layer. When lateral connections do exist, the activation landscape will be different. Third, it is assumed that during derivation, all input units other than  $i$  and  $j$  are never activated, in order to avoid the interference from other input units. All these limitations indicate that the results obtained here are only applicable to a model that is already significantly simplified. However, simulations show that even our complicated motor control model is regulated by these analytical results to some degree. For example, it was found in our simulations that increasing the magnitude of  $-c_s$  (while fixing  $c_p$  value) does effectively prevent the incoming weights of certain units from becoming equal during the training process.

In summary, the analysis results in this section, although having significant constraints on their applicability, give us some insight into the dynamics of our motor control model. These results provide a theoretical understanding about some of our empirical findings during simulations, and will give guidance help in building new network models in the future.

## Chapter 8

# Conclusion

### 8.1 Summary and Contributions

The research described in this dissertation developed a multi-layered, “closed-loop” motor control system with both sensory input and motor output. Most previous neural network motor control systems have used a single layer network that serves as both sensory input and motor output. These past models mostly used visual information only as sensory input, and did not examine map formation. The motor control model described in this dissertation is the first motor control model that incorporates proprioceptive sensory input, and it simulates map formation in primary proprioceptive cortex (roughly Brodmann area 3a and some surrounding cortex [Wise & Tanji, 1981]) using unsupervised learning method. This approach makes the model more biological realistic. Since visual input is an important sensory input, it was also implemented in the model. Simulations were conducted on models with either proprioceptive or visual input, as well as on models with combined proprioceptive and visual inputs.

The motor control model with only proprioceptive input was trained for the study of cortical feature map formation. It was found that, based on random initial stimulation in motor cortex, the network could self-organize to form cortical feature maps in both proprioceptive cortex and motor cortex. These maps forms clusters for specific features. The output maps found in primary motor cortex have characteristics that resembles those found in biological experiments [Donoghue *et al.*, 1992]. For example, elements that control the same muscle are widely distributed in motor cortex; many cortical motor neurons can activate multiple muscles. Moreover, the coexistence of multiple feature maps in the same cortical layer provided interesting relationships between each other. These maps formed meaningful alignment (or overlapping) after training, reflecting the mechanical constraints of the model arm. It was found that the length input map of a particular muscle was aligned with the tension input maps of its antagonist muscle. This occurred in both proprioceptive and motor cortex. Also in motor cortex, it was found that the motor output map of a particular muscle was aligned with the length input maps of its antagonist muscle and the tension map of its own. The study of these cortical feature map alignment provided us with better understanding of how sensory information was processed in the sensory and motor cortex and how this information came to influence the motor output. The work will be helpful in building future motor control models that incorporate more biological ingredients. This model has also become a substrate for lesion study in motor cortex [Goodall *et al.*, 1997], as illustrated in Appendix B.

We also examined the simulation of a version of the model which uses visual information instead of proprioceptive information as sensory input. Since the visual information is received by MI

via associative cortical areas in biological systems, and the actual encoding of visual information received by MI is unknown, three dimensional coordinates (with coarse coding in each dimension) of hand position was used as an abstraction of the visual information that reaches MI. Simulation results indicated that after training, visual input maps have formed in primary motor cortex. These maps also reflected the characteristics of arm mechanisms and formed meaningful input output relationships. Unlike the proprioceptive input maps, which are usually fully aligned with a particular motor output map, visual input maps are sometimes partially aligned with a motor output map. Moreover, each visual input map may be partially aligned with multiple motor output maps, and *vice versa*. A quantitative method was applied to measure how strongly these maps are aligned, or spatially correlated. It was found that the strength of such spatial correlation between certain pairs of input and output maps is related to the temporal correlation of both features during the training process.

After studying the motor control model with either proprioceptive and visual input alone, simulations were done with a version of the model with combined proprioceptive and visual inputs. The focus of this study was on how these two different sensory inputs interact in the cortical layer, and how their cortical feature maps influence each other. We wanted to know whether those characteristics of feature maps observed in separate models are still preserved in this combined model. It was found that the nature of proprioceptive input maps and their interrelationships are mostly unaffected, compared with the proprioception-only model. This indicates that adding visual input to the model does not influence the proprioceptive input maps significantly. However, the proprioceptive input did influence visual maps in some sense. Although the appearance of visual input maps remains similar with or without proprioceptive input, their relationship with motor output maps did change. Mostly, with the presence of proprioceptive input, a visual sensory input map could no longer be strongly correlated with the output maps of a pair of antagonist muscles at the same time. This indicates that adding proprioceptive inputs changed the relationships between visual input maps and the motor output maps, so that antagonist muscles are differentiated in visual input maps. In the meantime the coexisting of both proprioceptive input maps and visual input maps also formed consistent relationships that support the temporal correlation hypothesis described earlier.

Limited theoretical analysis on cortical map formation was also done. Due to the complicated structure and dynamics of the original motor control model, the analysis was necessarily based on a simplified model, with only one input layer and one output layer. First, the activation pattern of the output layer was analyzed based on single-input stimulation, in order to study the shape of feature maps when a specific input feature was present. Under the condition of absent lateral connections, the convergence of the activation pattern in the output layer could be proved under a competitive distribution of activation mechanism. The corresponding activation levels could be calculated. Subsequently, an analysis was done on why correlated (aligned) feature maps are usually related to the temporal correlation in input features during training. It was proved that perfectly correlated input features could produce fully aligned feature maps, with very few constraints. On the other hand, when two input features are in perfect anti-correlation, then there is no guarantee that the corresponding feature maps will be mutually exclusive. The correlation of the resulting feature maps depends on the parameter settings and weight conditions in this latter case. A sufficient condition of forming mutually exclusive feature maps was derived.

In summary, the research described in this dissertation has developed a biologically plausible neural network motor control system. The system self-organized via training to form multiple input

and output cortical feature maps. These maps exhibited some properties that are consistent with experimental finds in biological systems, such as distributed feature representation, controlling of multiple muscles for individual MI neurons, etc. The relationships between cortical feature maps reflect the temporal correlation hypothesis: *temporally correlated features usually cause their corresponding cortical map representations to be spatially correlated*. This hypothesis is strongly supported by simulation results in different versions of the motor control model. It is also supported by theoretical investigation on a simplified model, and provides testable predictions that can guide future experimental research.

## 8.2 Limitations and Future Work

The motor control model described in this dissertation provides useful information about the organization of cortical feature maps. This model, however, has its limitations. As a motor control model, it is a great simplification from biological motor control systems. This model tried to simulate sensory and motor cortices in a realistic fashion, but the lack of detailed biological data in cerebral cortex and the possible computational cost have both imposed limitations on this effort. The simulation of the model arm is even further simplified, as the emphasis was put on the study of cortical feature maps. Although hundreds of elements were contained in the cortical layers, very few elements were used to simulate arm neurons. For each muscle group, there was only one element representing the activation of the entire group of neurons controlling this muscle group. Such an simplification, although enough in supplying sensory information to cortical layers, is far from biological system, where each muscle is controlled by many motor neurons.

This motor control model was developed for the study of cortical feature maps, and not for actual arm position purposes. The unsupervised learning method used in the model did not force the movement of the arm towards any specific target. As a result, the model is not able to perform target reaching tasks accurately. However, during training, the stabilized network caused some kind of association between visual input and motor output, as described earlier. So this model did exhibit some improvement in moving the arm close to a position where corresponding visual input is supplied to cortical layers. But the overall performance of this model is incomparable to those motor control models built for industrial reaching task purposes, which do not care about biological plausibility.

The theoretical analysis in Chapter 7 was based on a simplified model. Therefore, the results obtained in this chapter serves as an illustration instead of a proof for the original motor control model. Some part of the analysis, such as the derivation of conditions under which the two anti-correlated input features cause mutually exclusive feature maps, has incorporated many assumptions, which limit the applicability of the results obtained. These analysis results could only help us understand better about the behavior of such type of network in general, and provide a helpful hint for empirically choosing experimental parameters. They are not mathematically applicable to our motor control model directly, even though many phenomena are similar.

Some future work in this motor control model may include building a more realistic arm model, with more elements representing muscle activation. Instead of using accurate activation of single elements to decide the degree of muscle contraction, different amounts of activated elements could be used to determine how much a muscle contracts. This approach is more similar to biological muscle control, where muscle activation is determined by the number of recruited motor units. Also the training methods could be changed in some fashion. Target reaching is an easy task for human

beings because we have an intention to reach that target. The current motor control model did not implement such an “intention”. It is possible that a reinforcement learn method could produce a better result. Moreover, additional mathematical analysis is necessary for this complicated motor control model. For example, one might study further how cortical maps are influenced by parameter settings. One might also want to study the stablization of activation patterns under competitive distribution of activation. Current simulations showed that the activation of the network always stablized after a certain number of time steps. This may be due to the right choice of parameters. But more likely it is a property of such kinds of networks in general. Some previous work has proved that total amount of activation in the network could converge under competitive distribution of activation [Reggia & Edwards, 1990]. But there is no guarantee that individual elements could reach an equilibrium point. In our work, the convergence for individual elements could be proved only when there was a single input stimulation and there were no lateral connections. So the proof of such convergence in a more general situation is of great interest. Similar to these above examples, there are still many phenomena about the behavior of the network that were consistently observed in simulations. These phenomena are also subject to theoretical investigation.



## **Appendix A**

### **Complete Cortical Feature Maps**

In this Appendix, cortical feature maps in different versions of our motor control models are listed. These maps are cataloged here to document representative examples of the model's behavior. In the model with proprioceptive input only, proprioceptive input maps in both PI and MI layer, as well as motor output maps, are listed. In the model with visual input only, both MI motor output maps and visual input maps are listed. In the model with combined proprioceptive and visual inputs, the listed maps are: proprioceptive input maps in PI and MI; motor output maps in MI; and visual input maps in MI.

#### **A.1 Cortical Feature Maps of Motor Control Model with Proprioceptive Input Only**

##### **A.1.1 Proprioceptive input maps in PI layer**

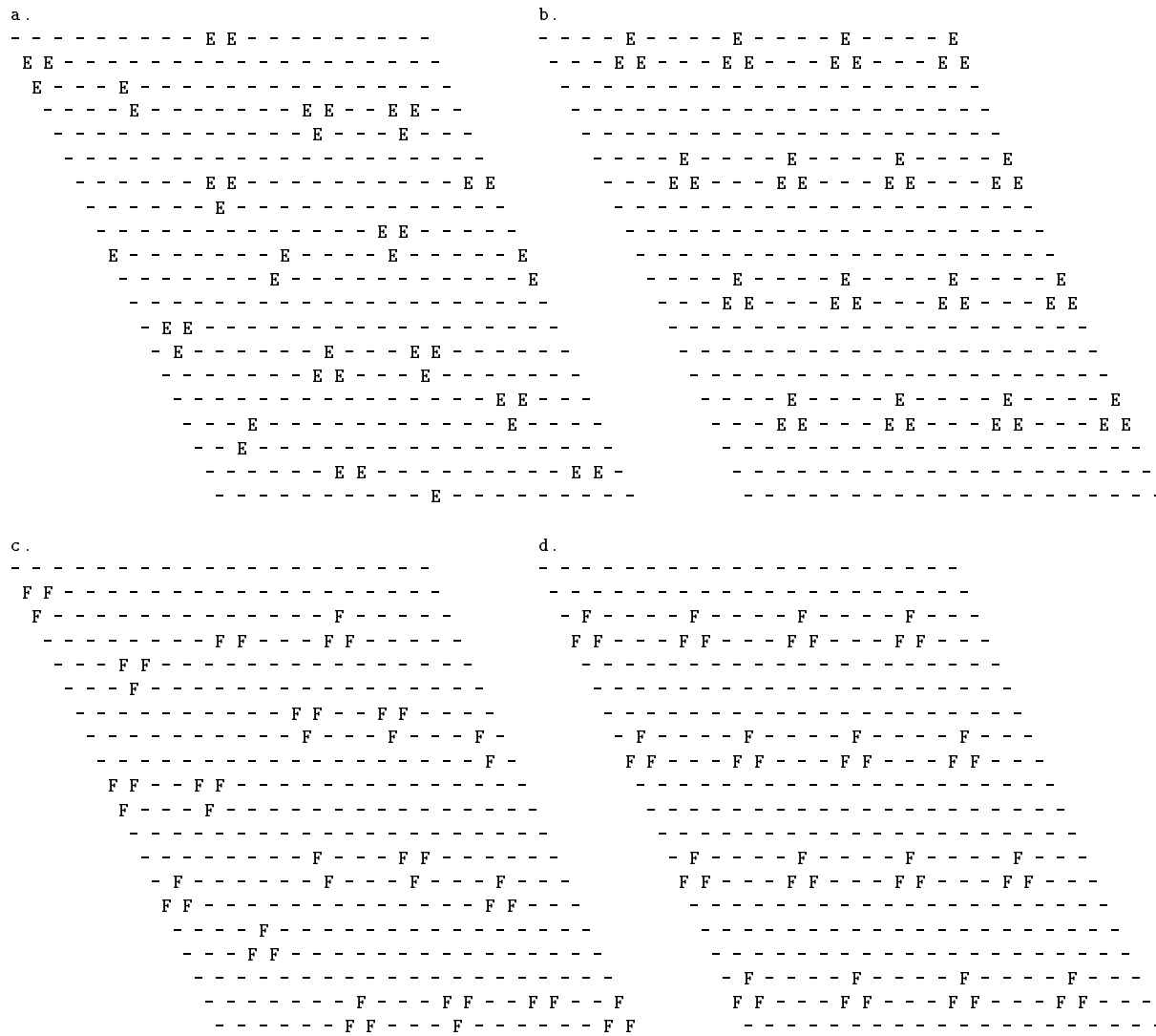


Figure A.1: Tuning of PI elements to the length of the upper arm extensor (E) and flexor (F) before (left) and after (right) training (threshold=0.4).

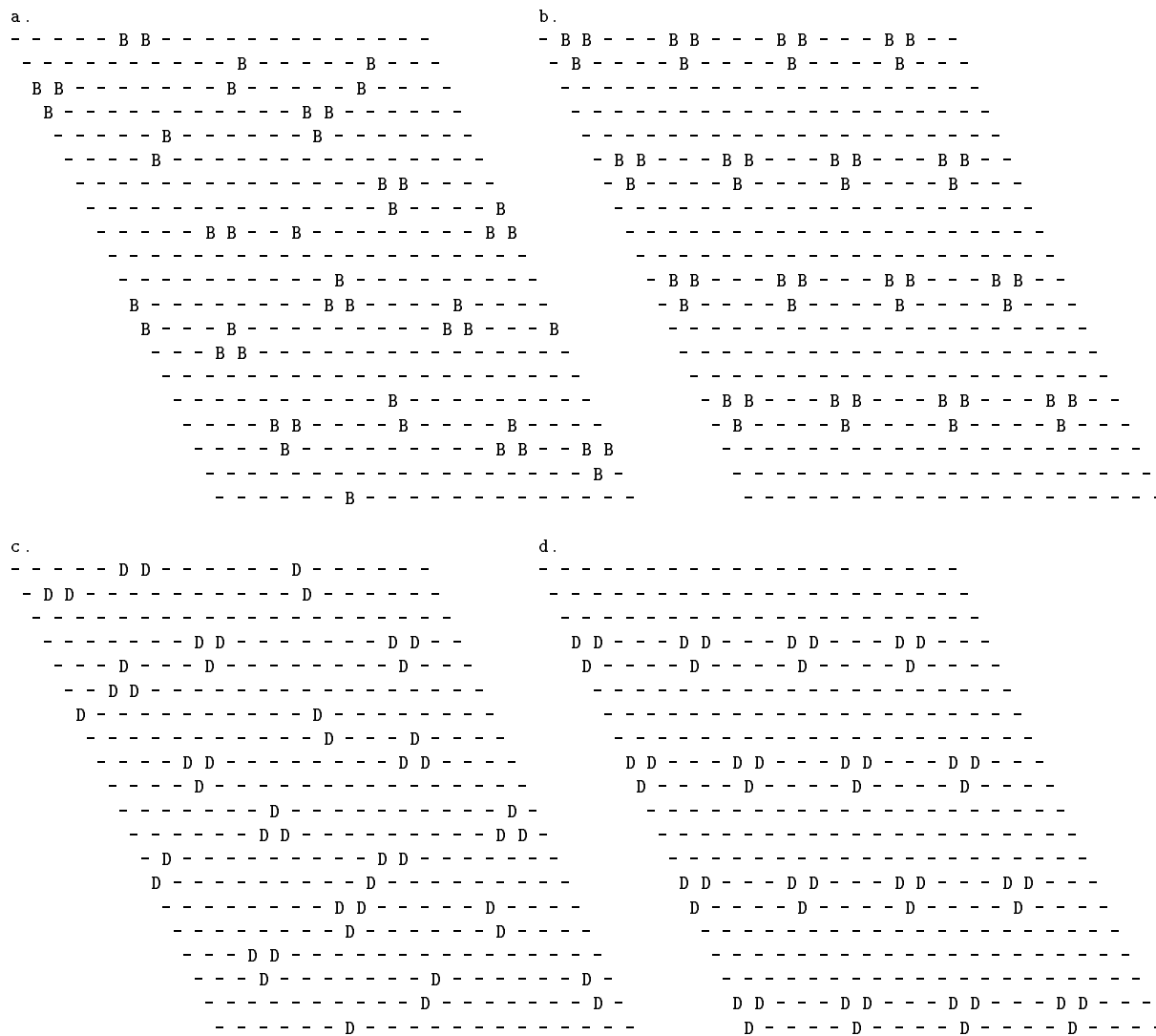


Figure A.2: Tuning of PI elements to the length of the upper arm abductor (B) and adductor (D) before (left) and after (right) training (threshold=0.4).

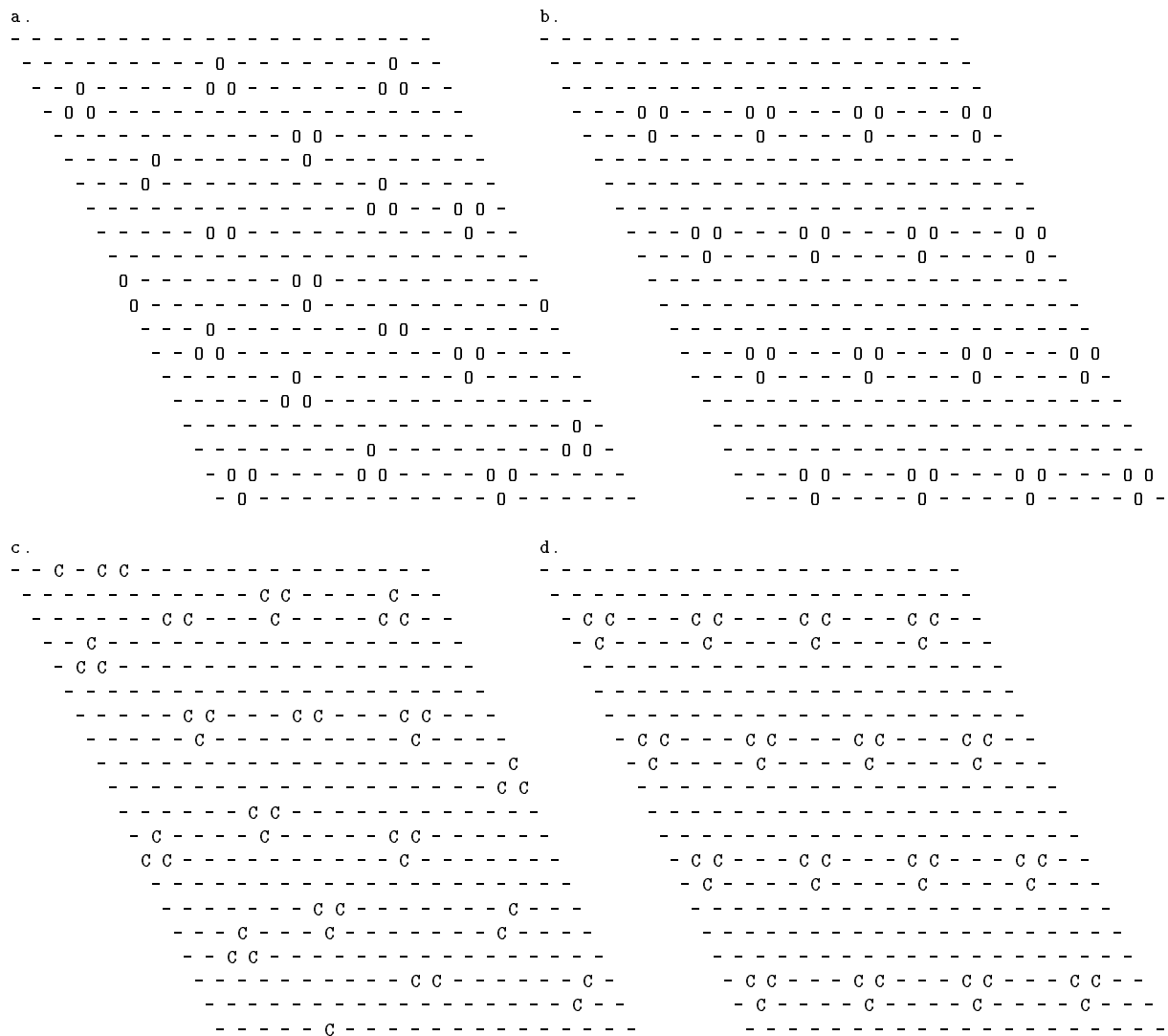


Figure A.3: Tuning of PI elements to the length of the lower arm extensor or operner (O) and flexor or closer (C) before (left) and after (right) training (threshold=0.4).

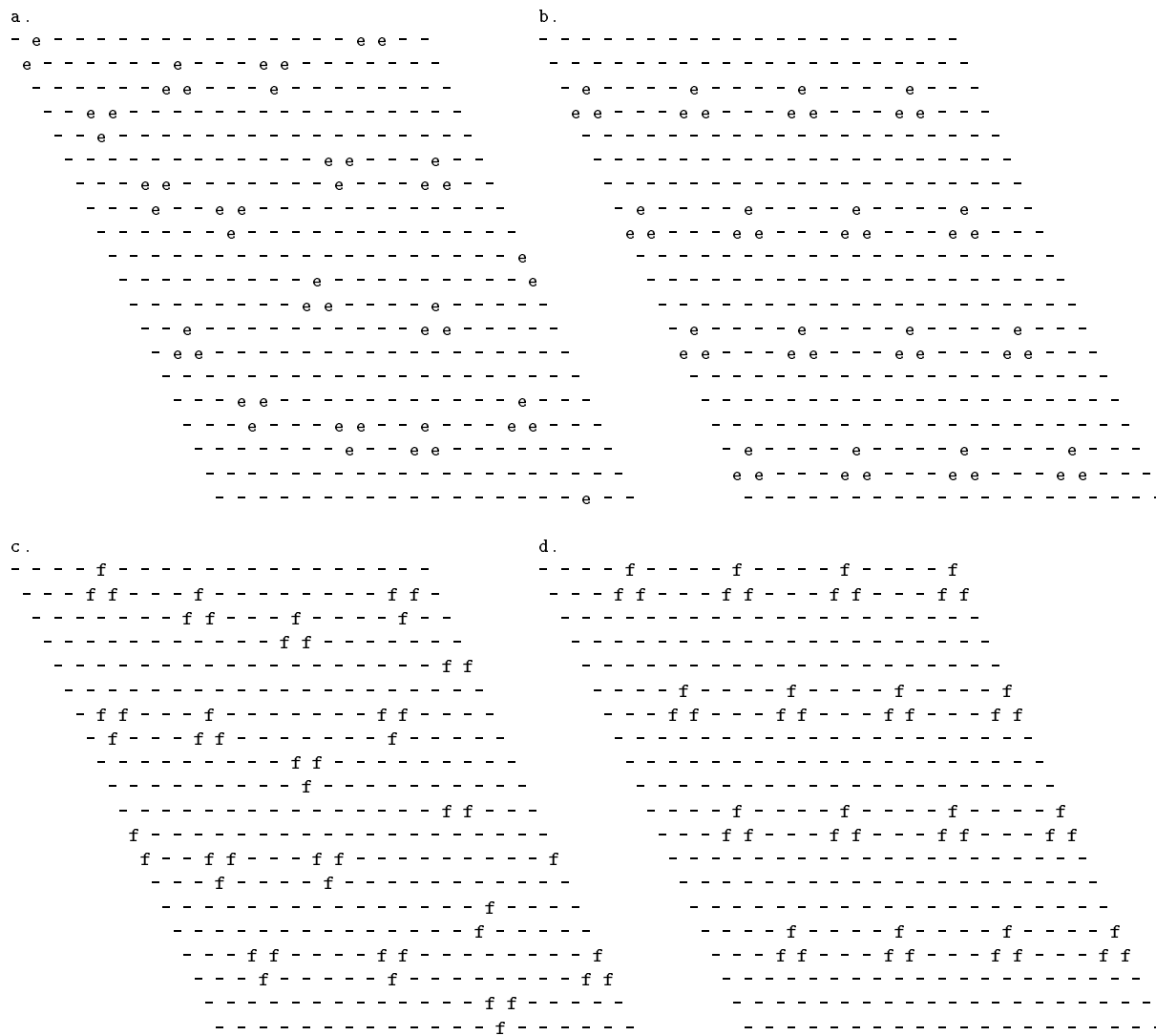


Figure A.4: Tuning of PI elements to the tension of the upper arm extensor (e) and flexor (f) before (left) and after (right) training (threshold=0.4).

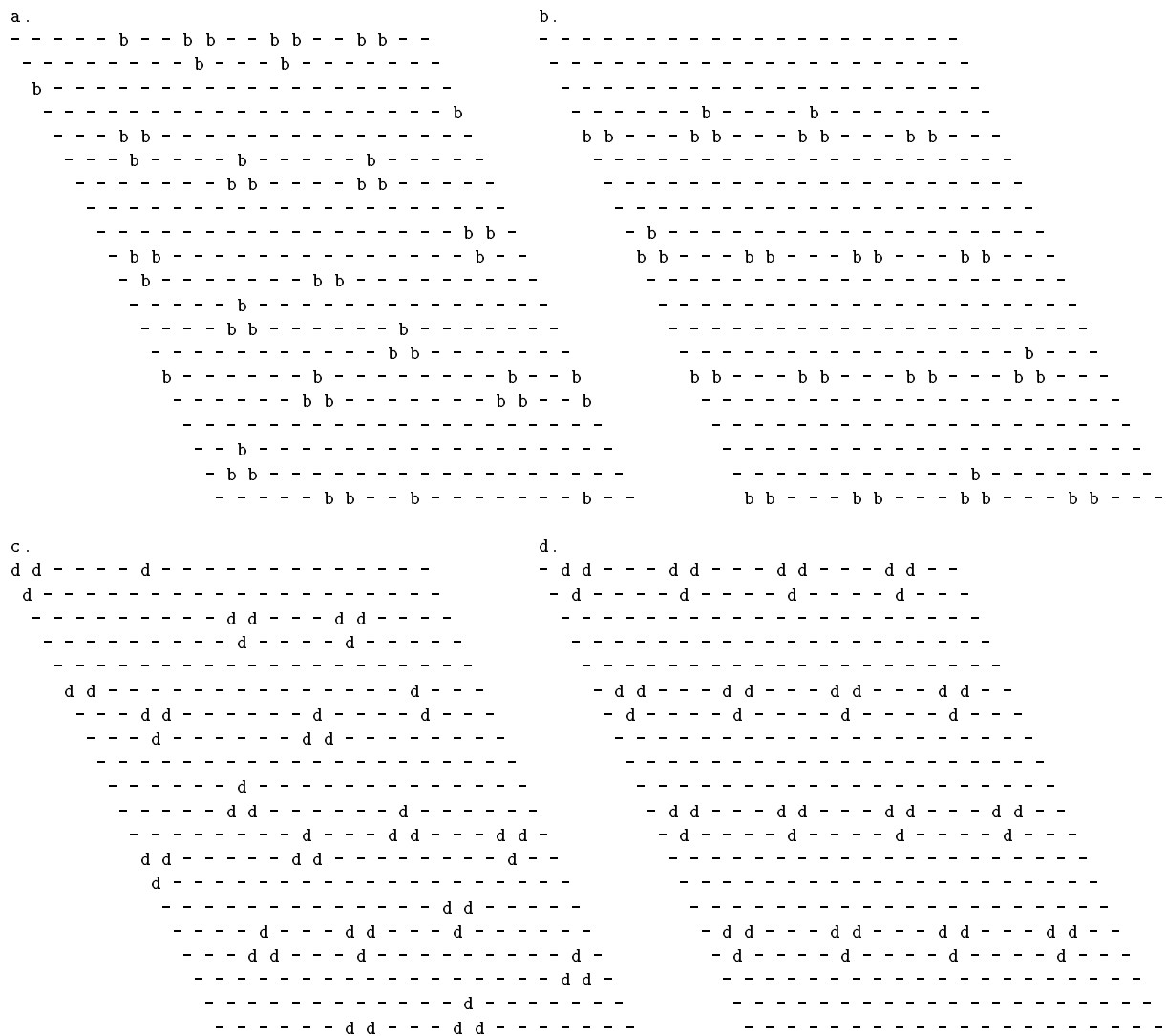


Figure A.5: Tuning of PI elements to the tension of the upper arm abductor (b) and adductor (d) before (left) and after (right) training (threshold=0.4).

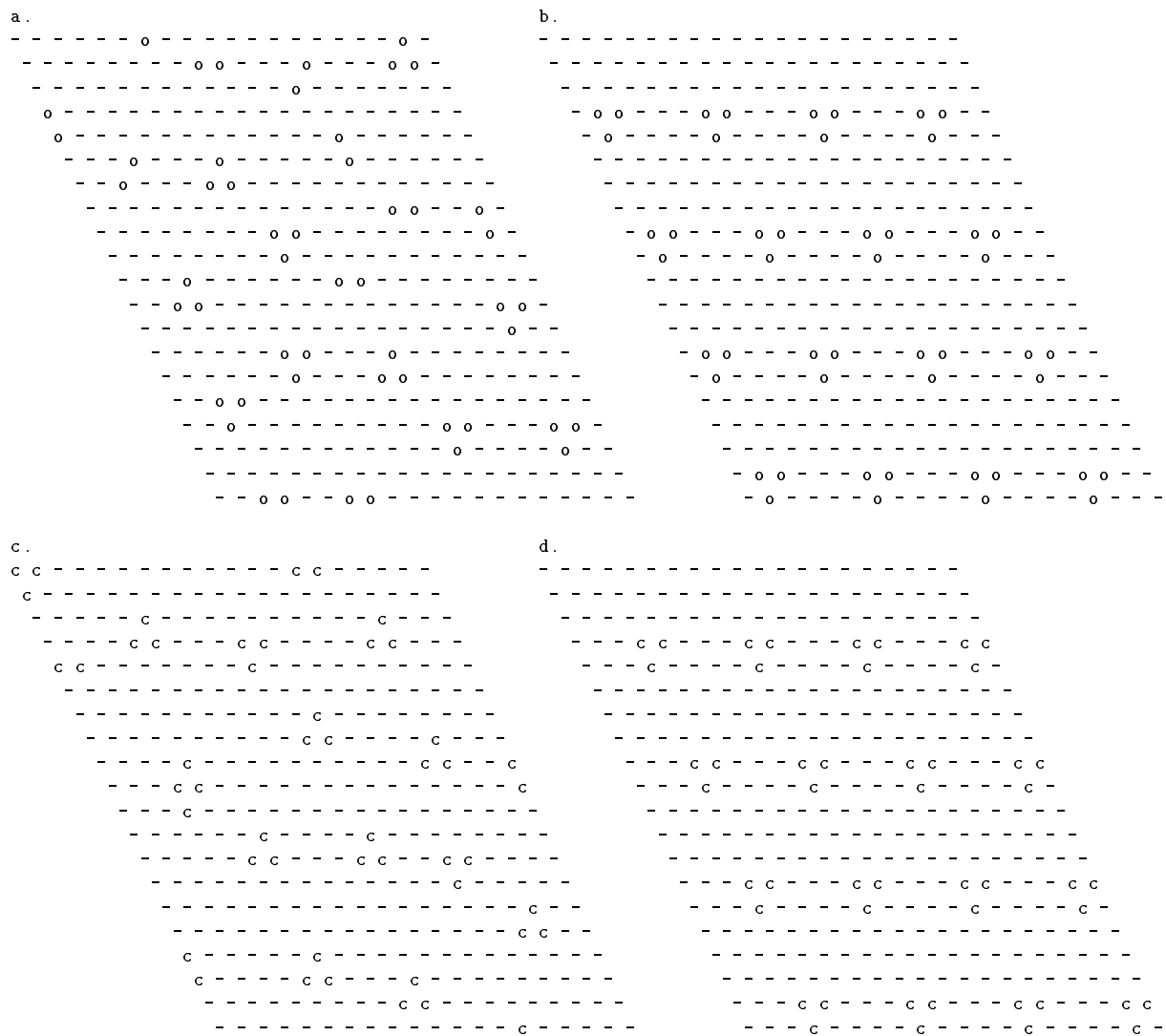


Figure A.6: Tuning of PI elements to the tension of the lower arm extensor or opener (o) and flexor or closer (c) before (left) and after (right) training (threshold=0.4).

### A.1.2 Proprioceptive input maps in MI layer

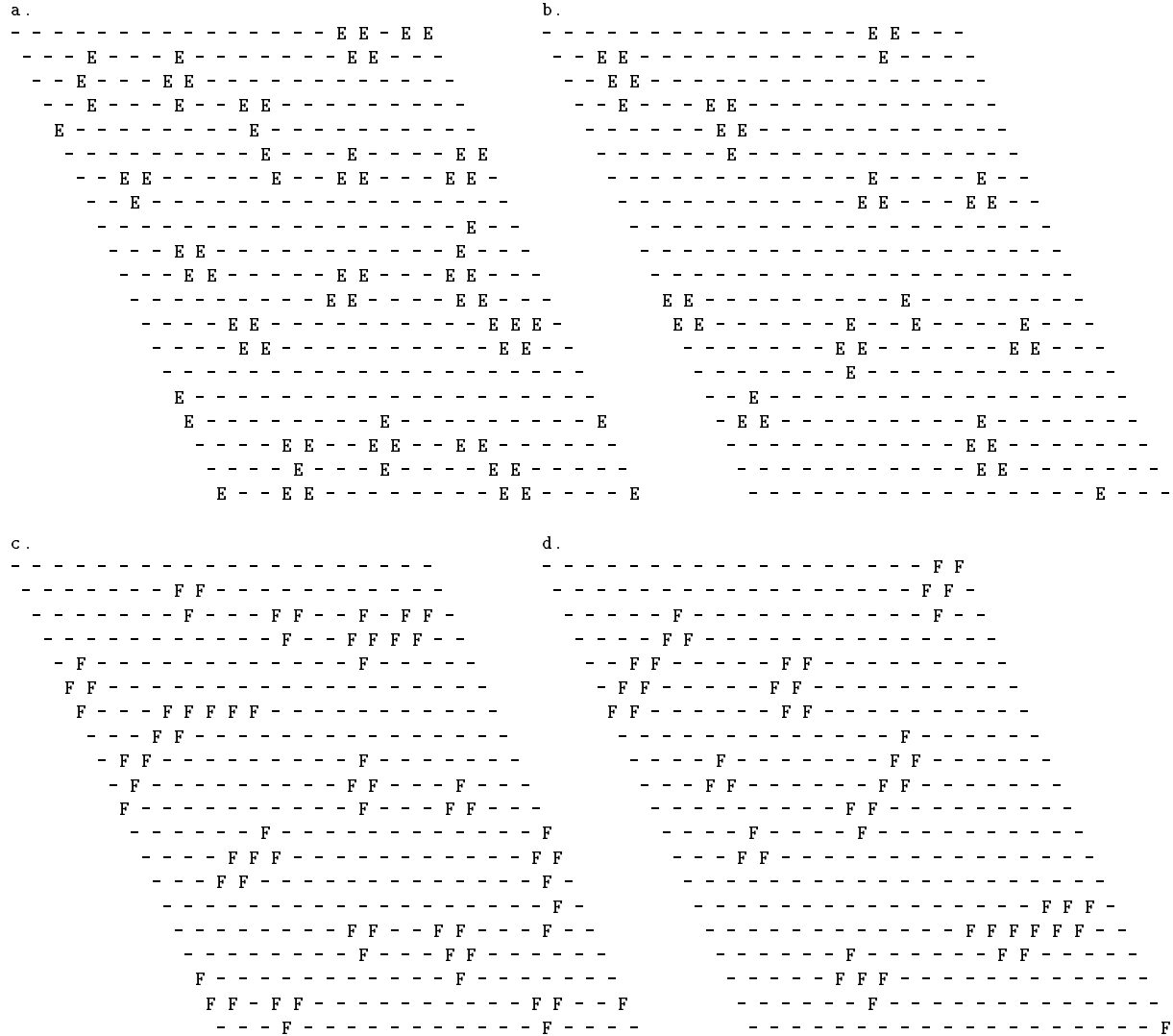


Figure A.7: Tuning of MI elements to the length of the upper arm extensor (E) and flexor (F) before (left) and after (right) training (threshold=0.4).



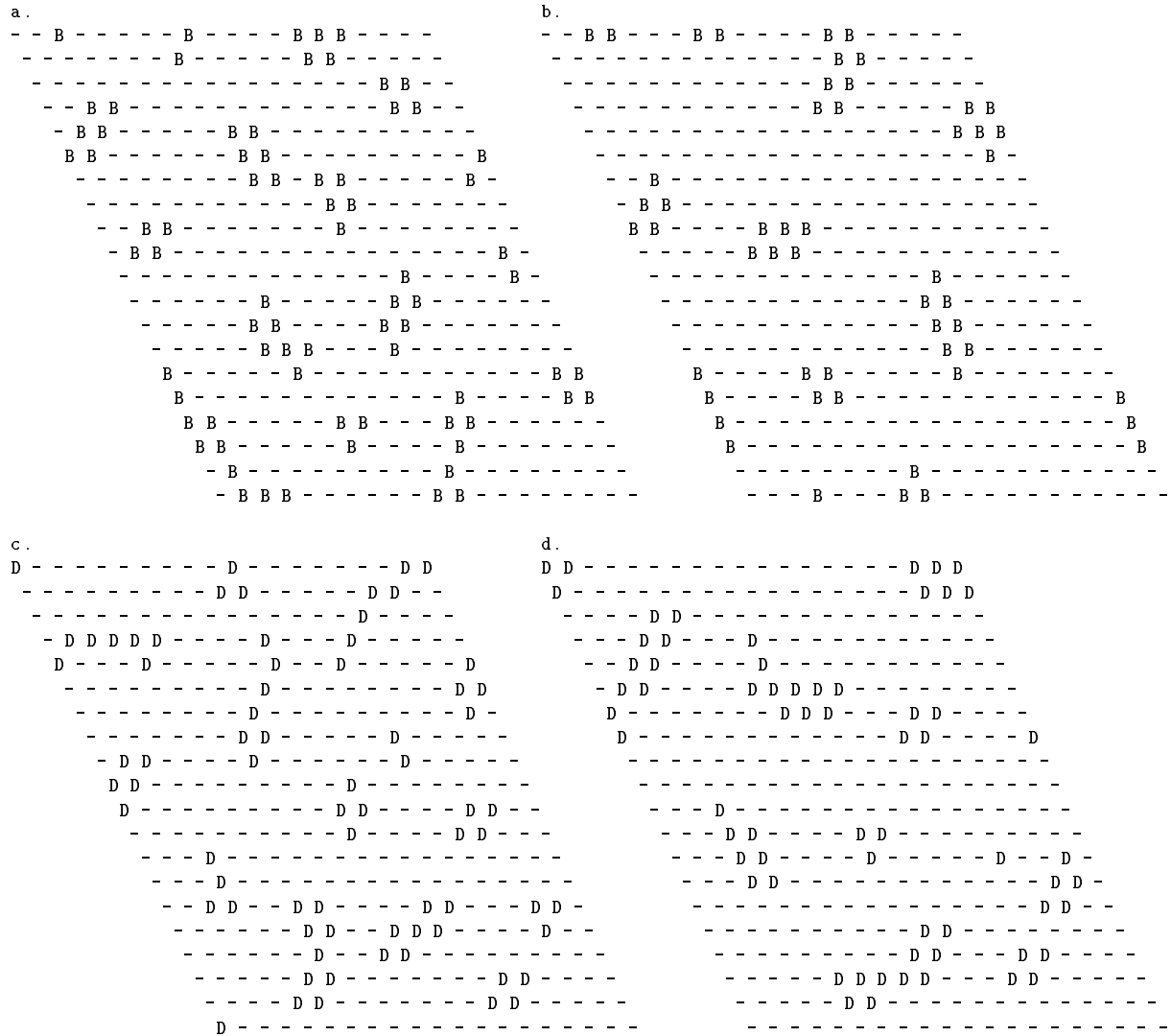


Figure A.8: Tuning of MI elements to the length of the upper arm abductor (B) and adductor (D) before (left) and after (right) training (threshold=0.4).

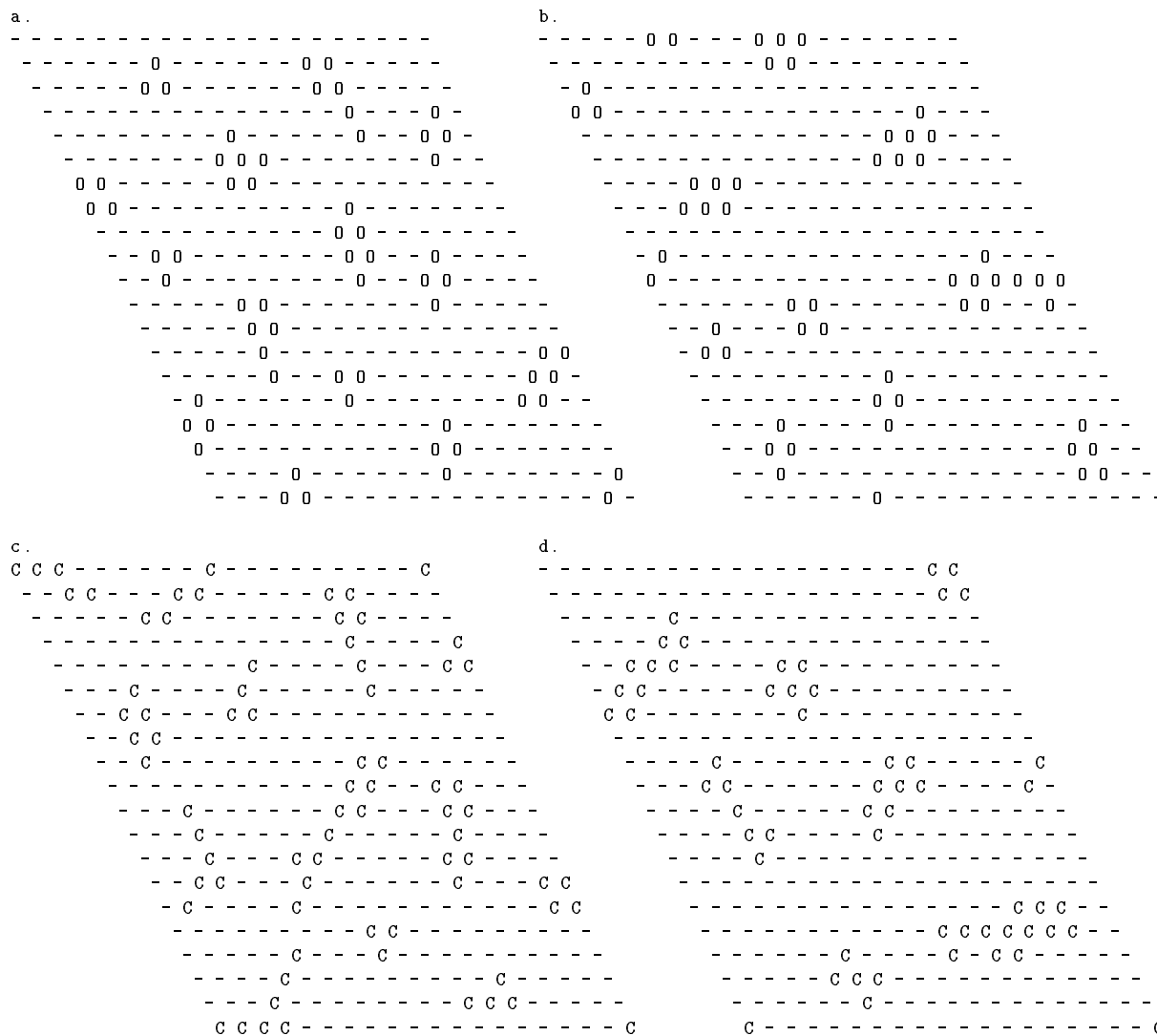


Figure A.9: Tuning of MI elements to the length of the lower arm extensor or operner (O) and flexor or closer (C) before (left) and after (right) training (threshold=0.4).

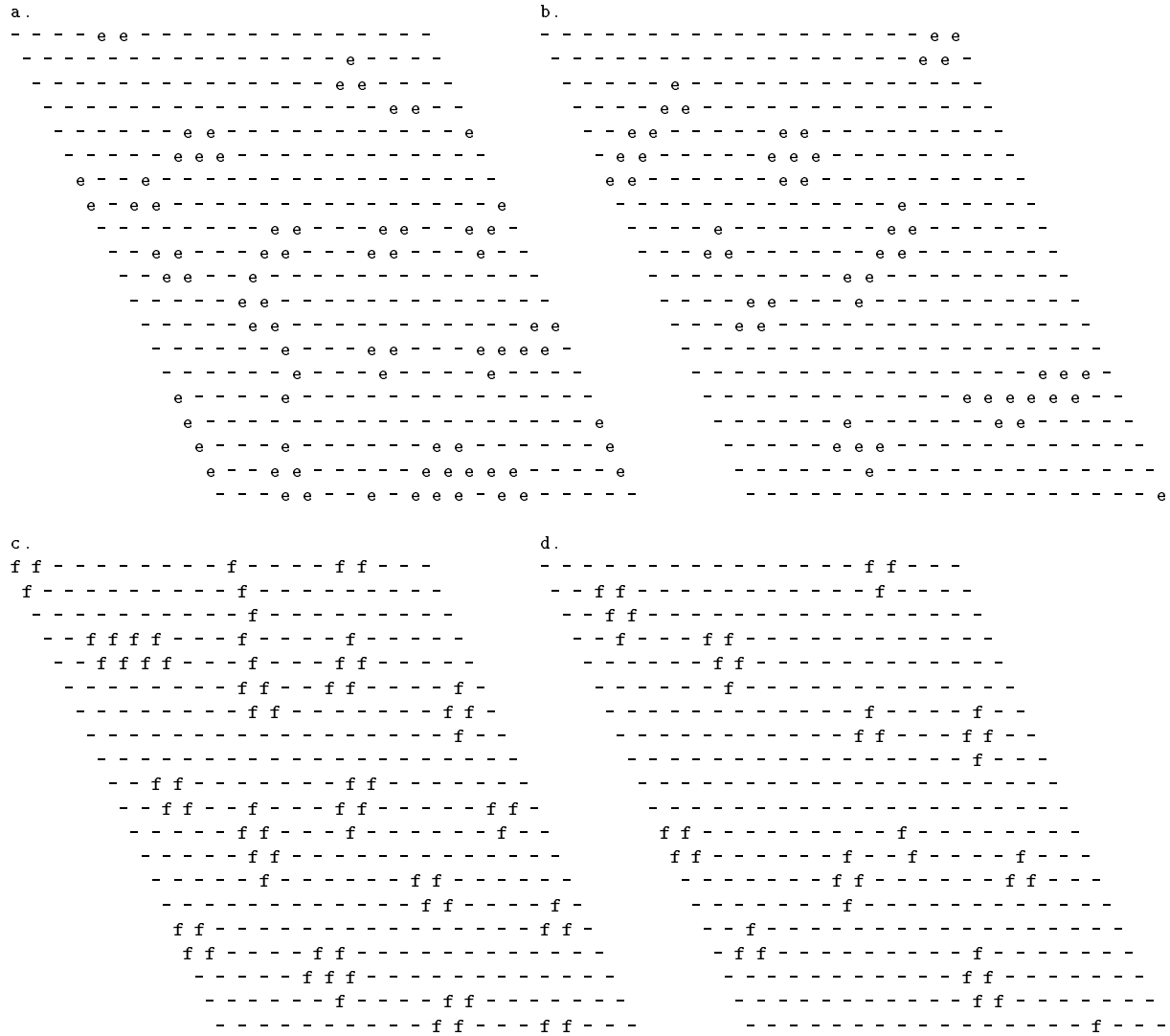


Figure A.10: Tuning of MI elements to the tension of the upper arm extensor (e) and flexor (f) before (left) and after (right) training (threshold=0.4).

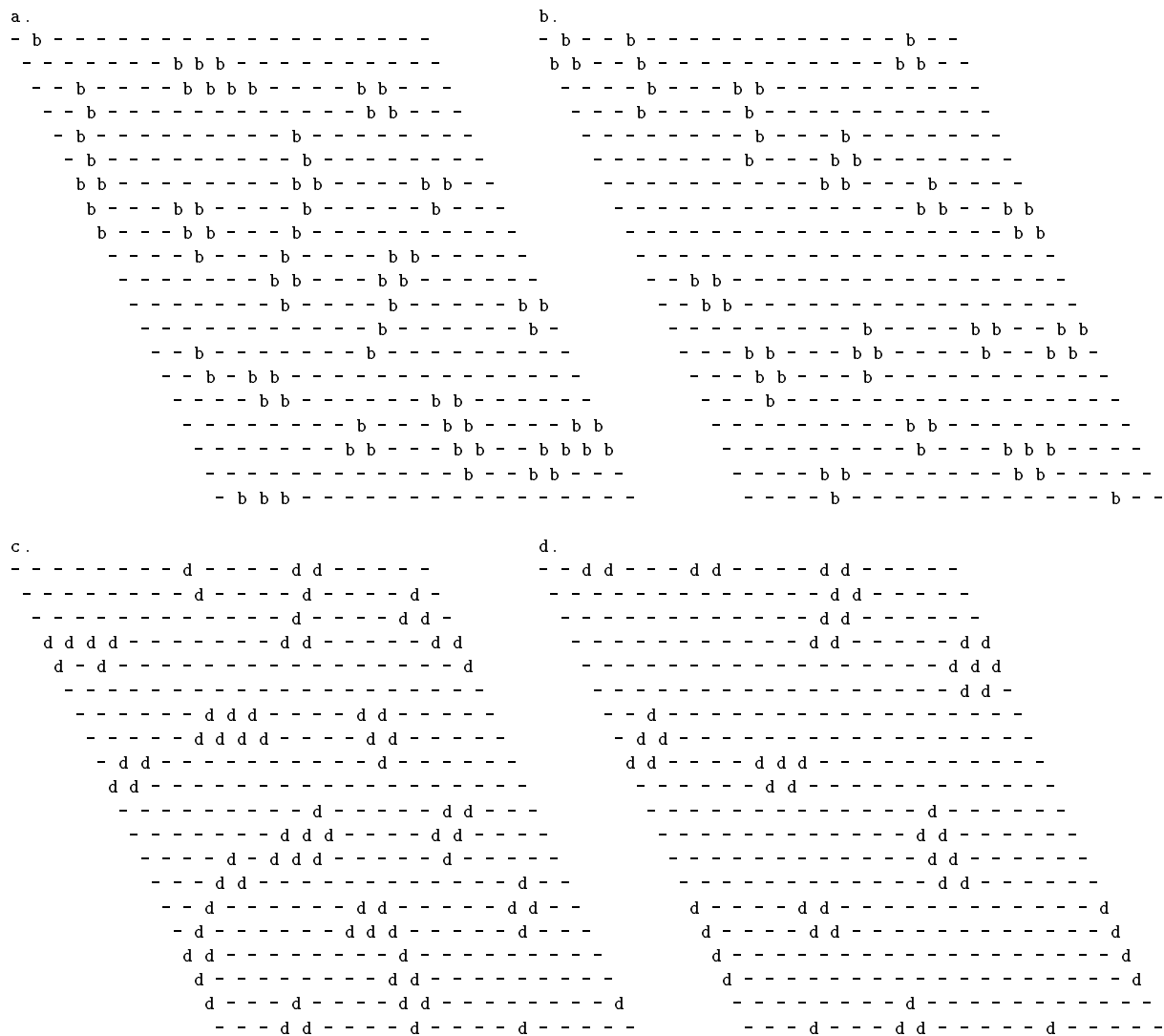


Figure A.11: Tuning of MI elements to the tension of the upper arm abductor (b) and adductor (d) before (left) and after (right) training (threshold=0.4).

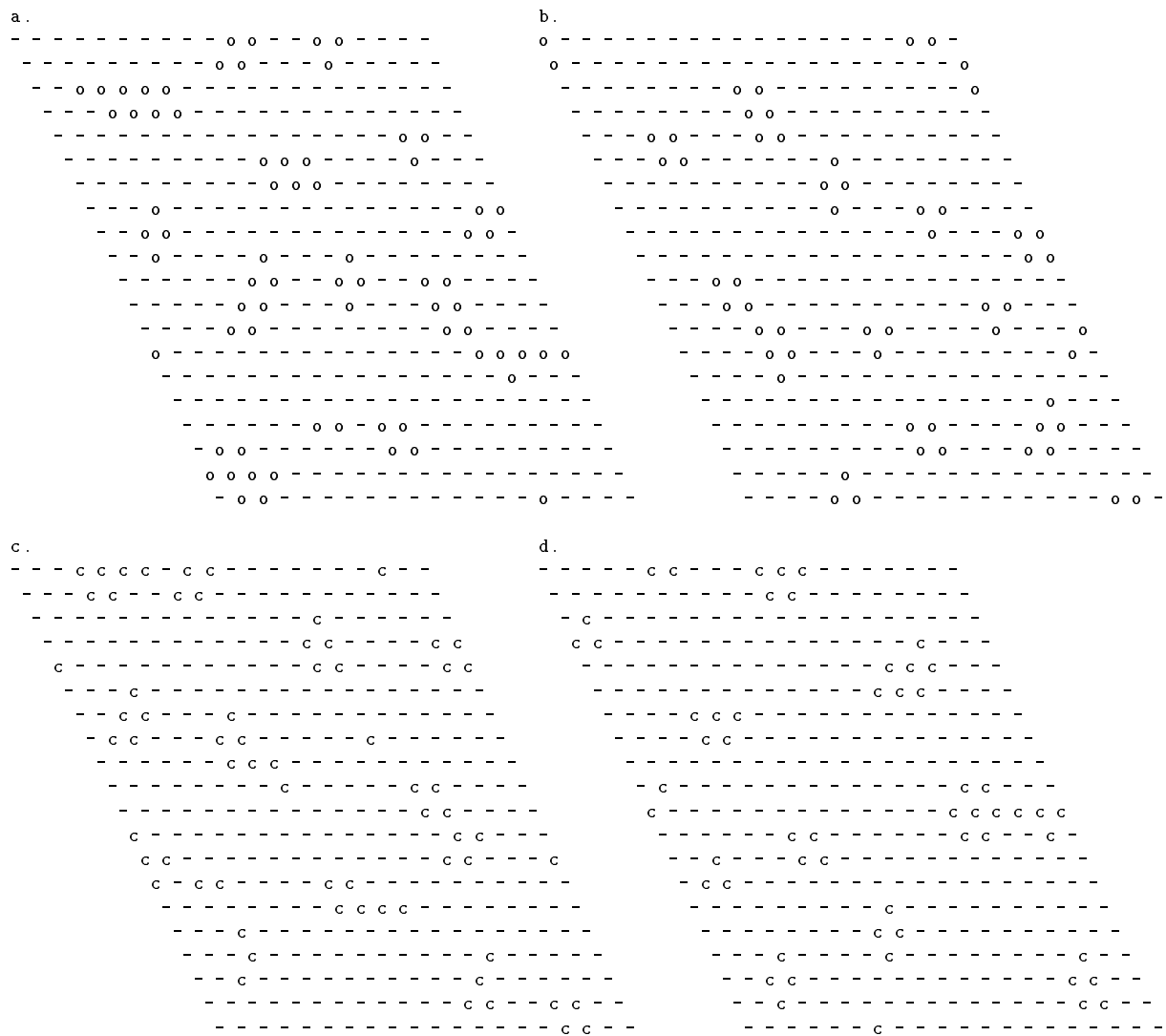


Figure A.12: Tuning of MI elements to the tension of the lower arm extensor or opener (o) and flexor or closer (c) before (left) and after (right) training (threshold=0.4).

### A.1.3 Motor output maps in MI layer

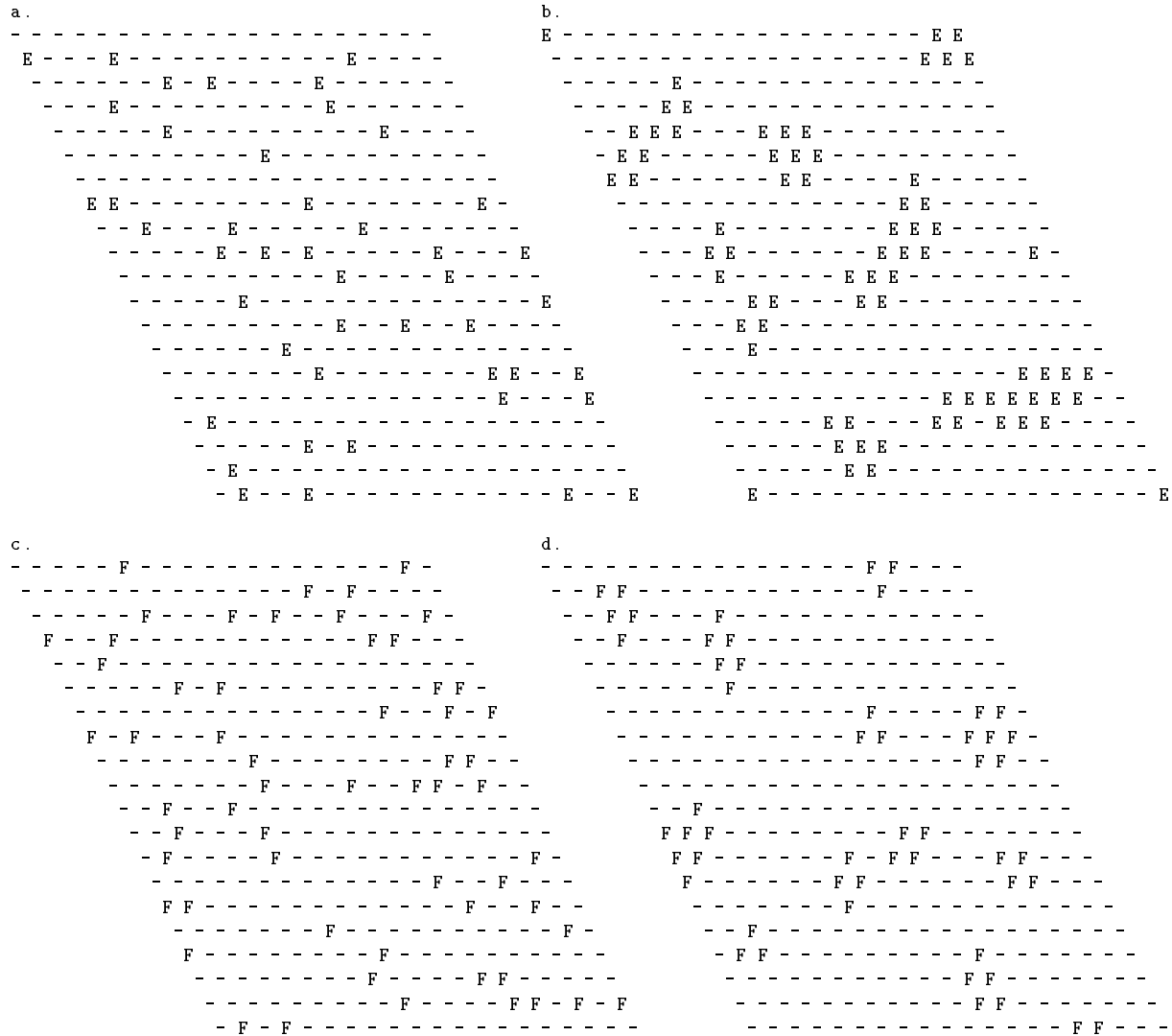


Figure A.13: MI output map before (left) and after (right) training for upper arm extensor (E) and flexor (F) (threshold=0.4).

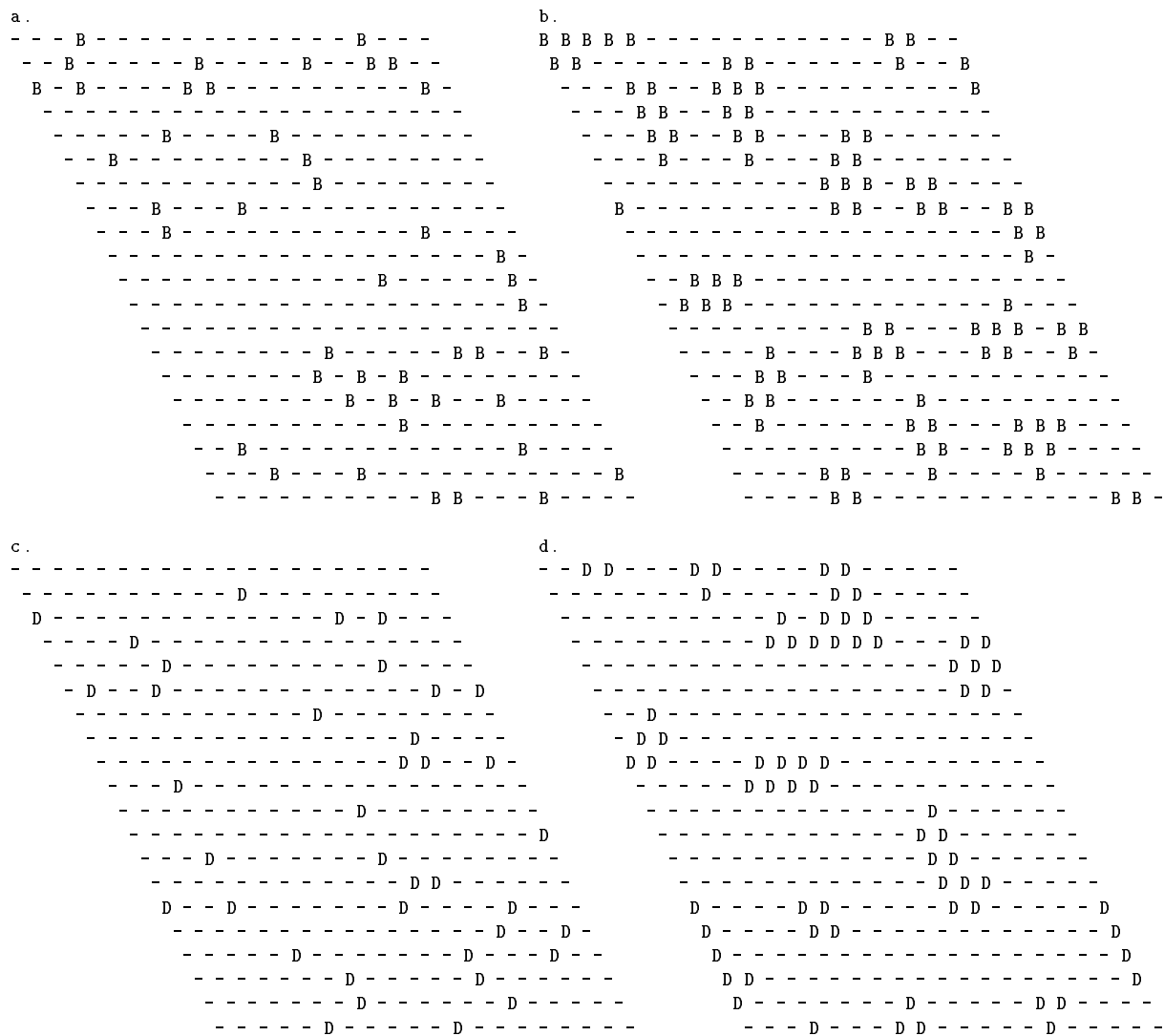


Figure A.14: MI output map before (left) and after (right) training for upper arm abductor (B) and adductor (D) (threshold=0.4).

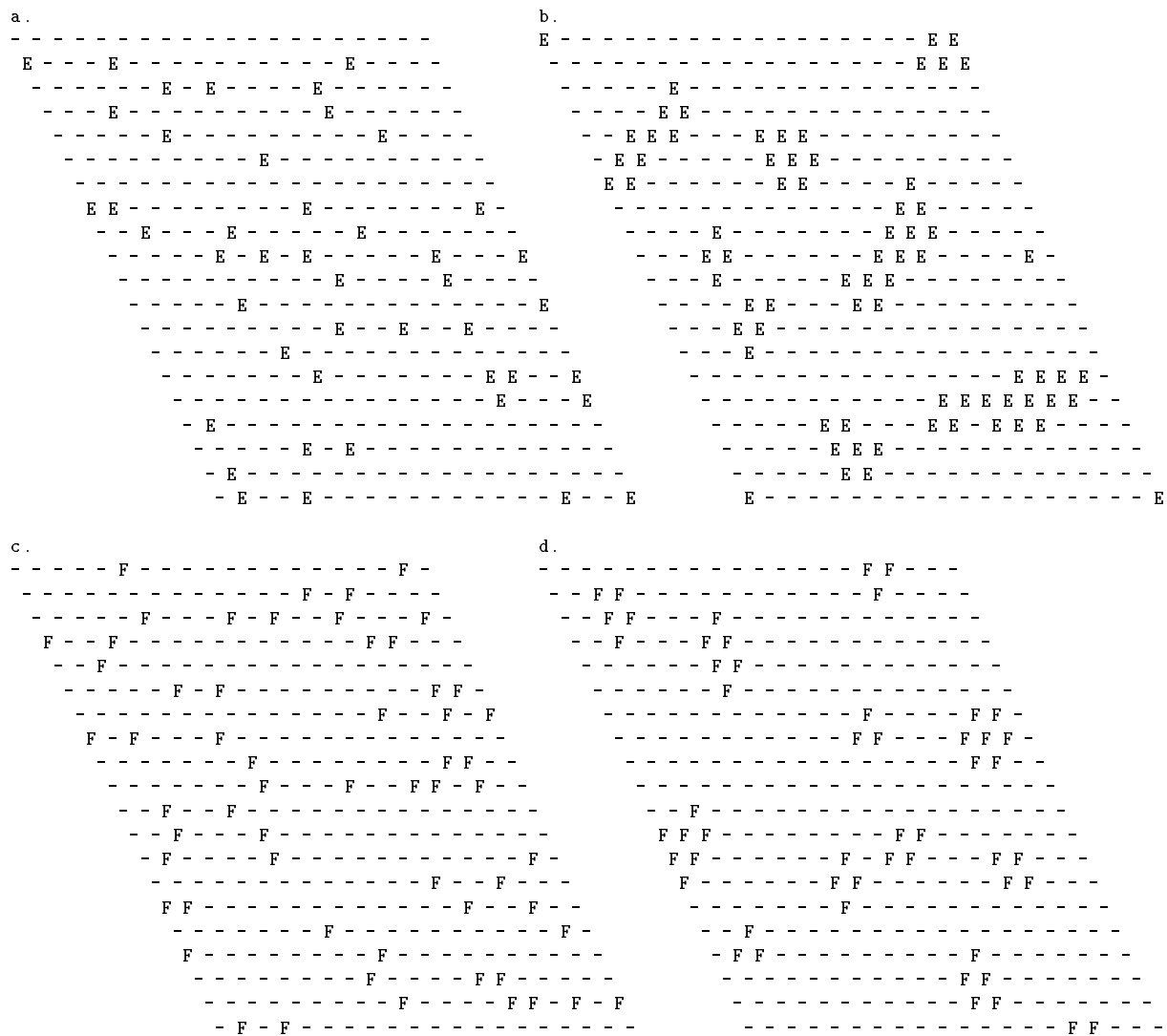


Figure A.15: MI output map before (left) and after (right) training for lower arm extensor or opener (O) and flexor or closer (C) (threshold=0.4).



## A.2 Cortical Feature Maps of Motor Control Model with Visual Input Only

### A.2.1 Motor output maps in MI layer

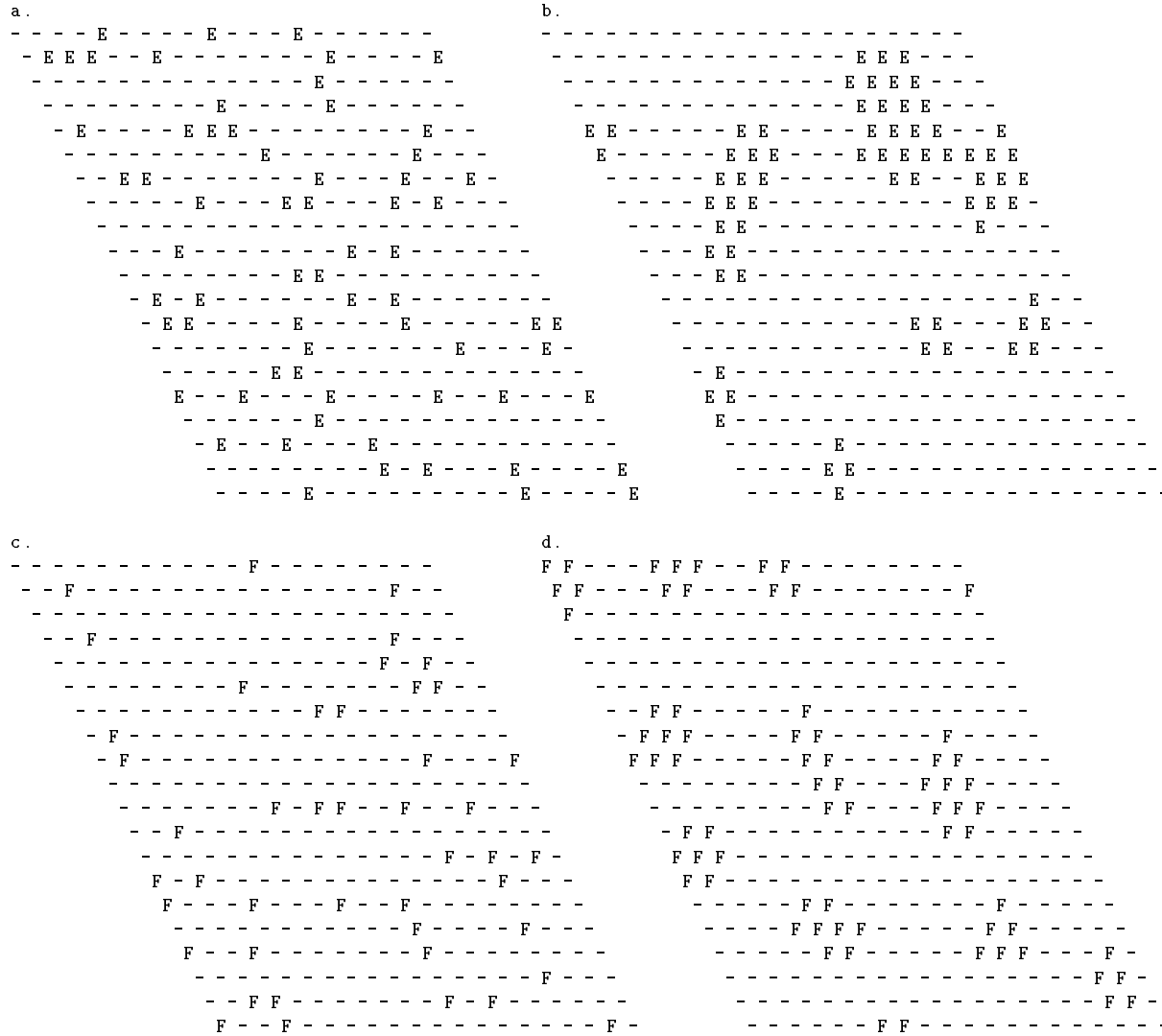


Figure A.16: MI output map before (left) and after (right) training for upper arm extensor (E) and flexor (F) (threshold=0.4).

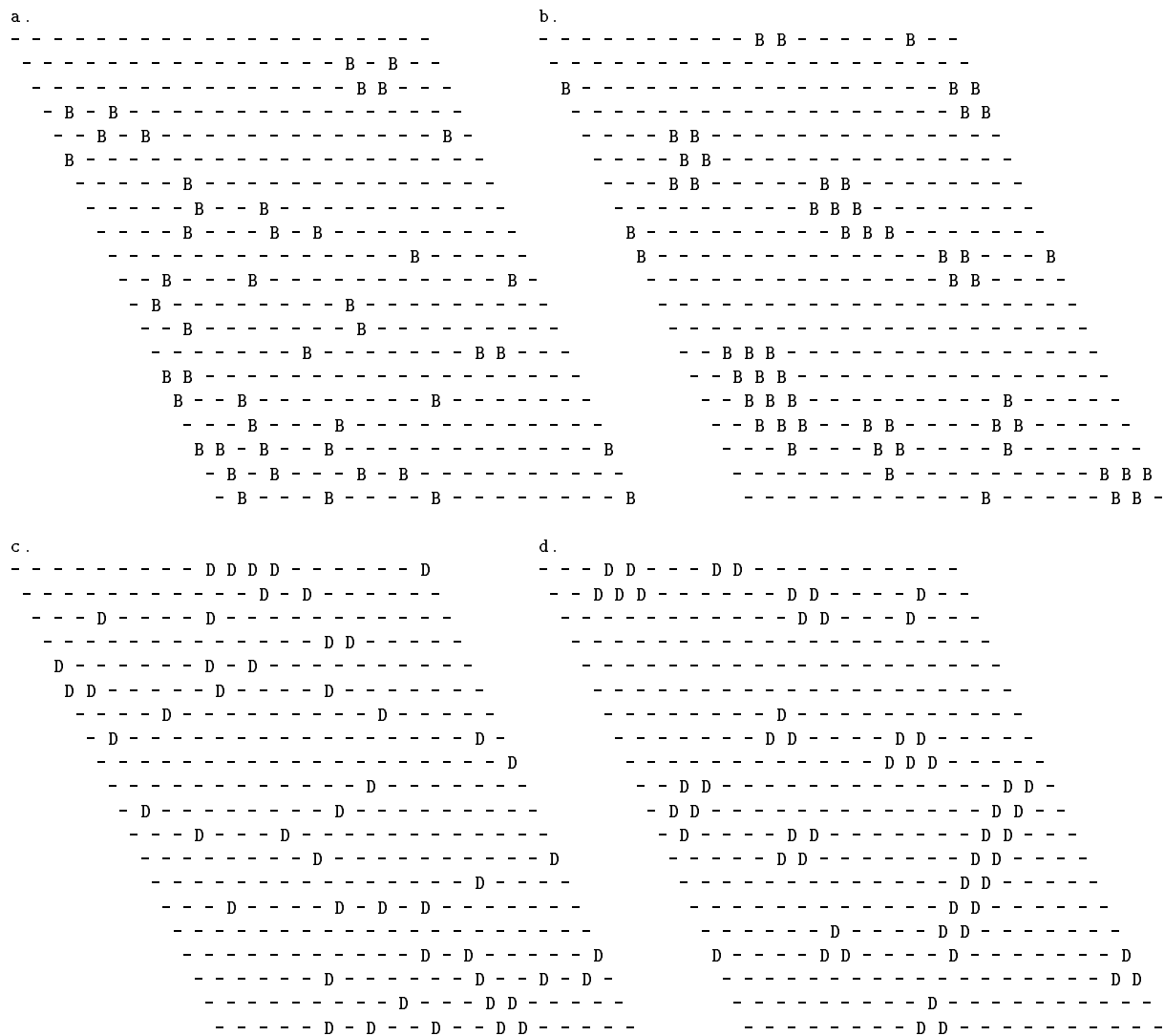


Figure A.17: MI output map before (left) and after (right) training for upper arm abductor (B) and adductor (D) (threshold=0.4).

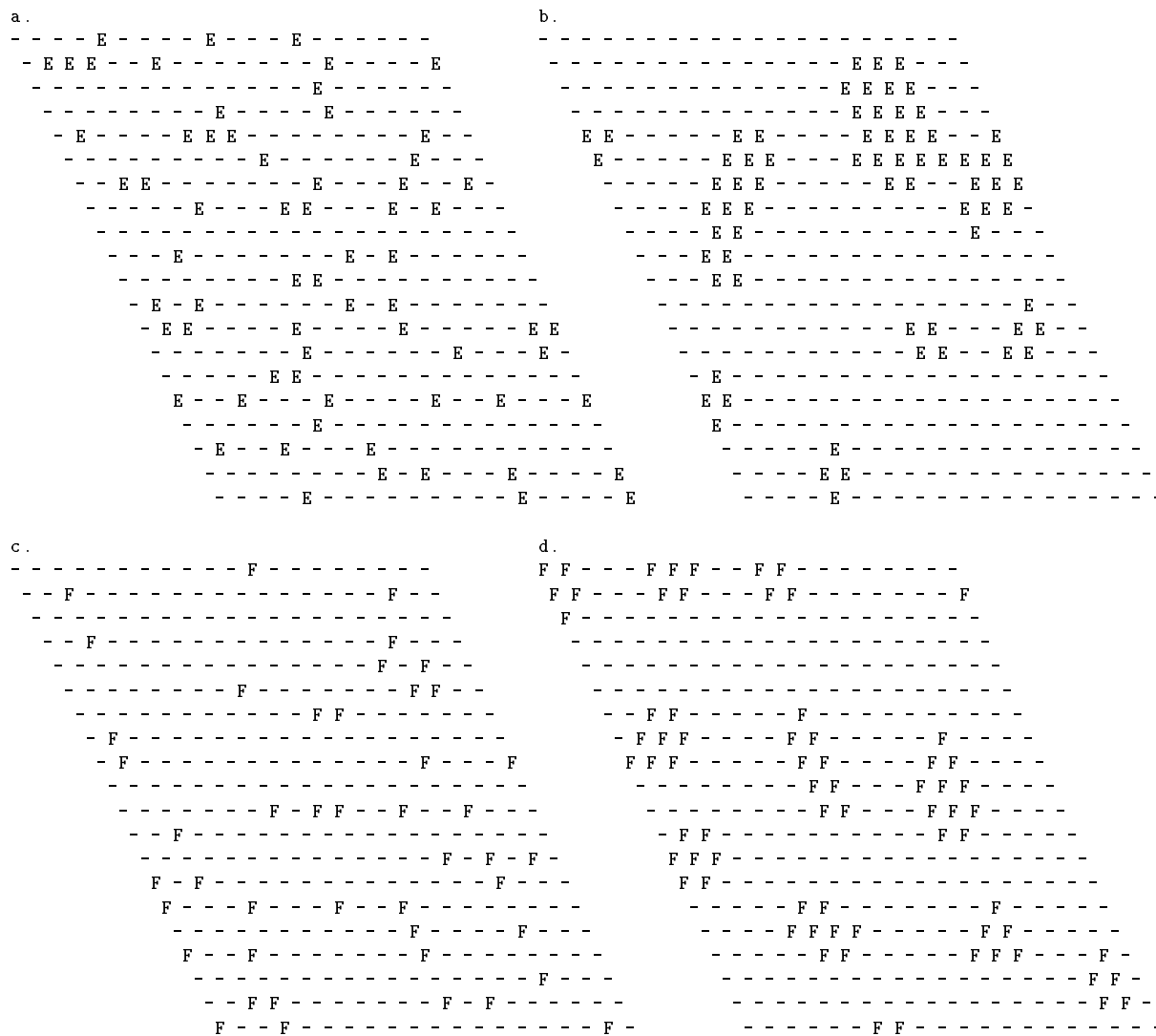


Figure A.18: MI output map before (left) and after (right) training for lower arm extensor or opener (O) and flexor or closer (C) (threshold=0.4).

## A.2.2 Visual input maps in MI layer

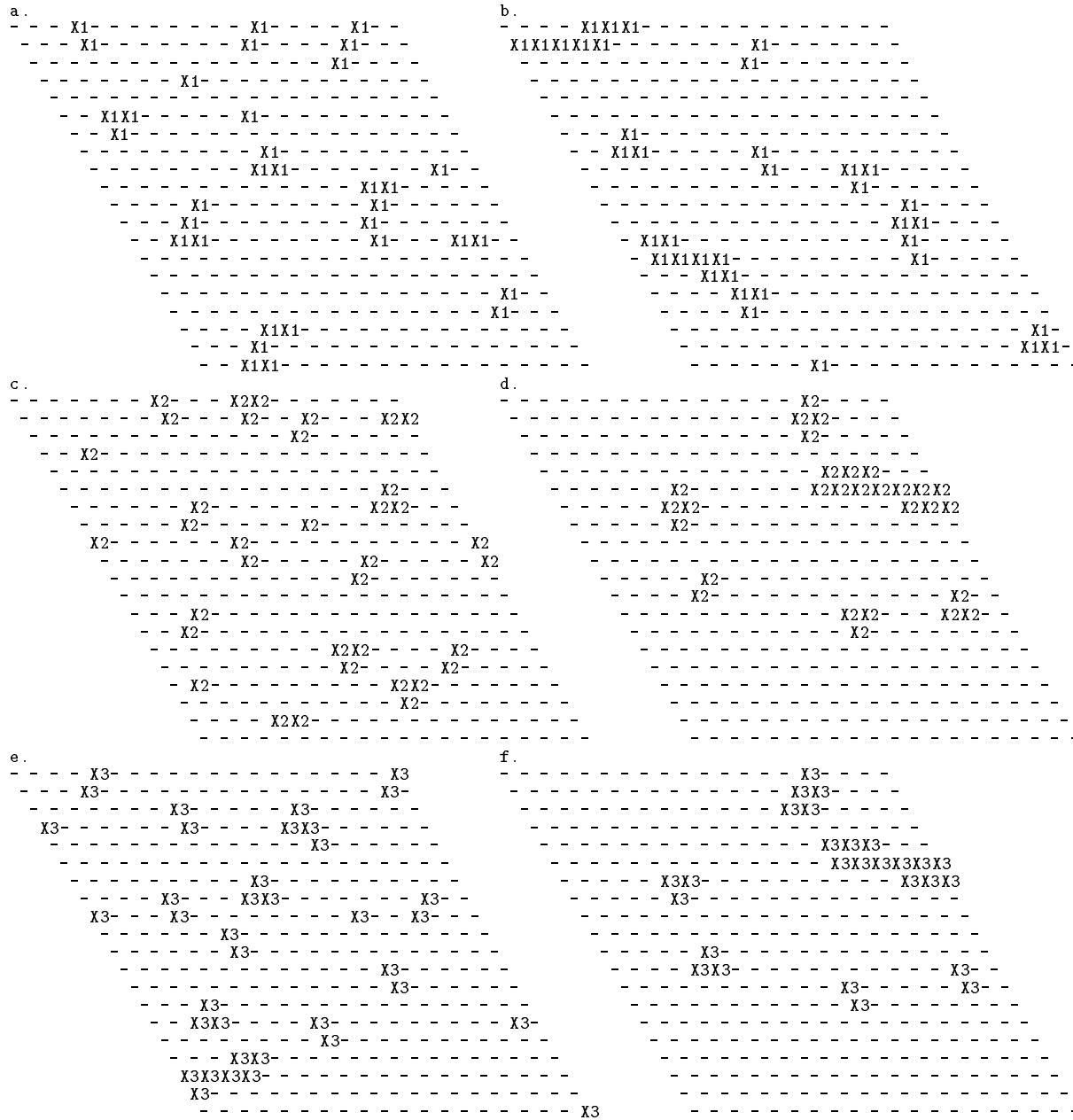


Figure A.19: The MI input maps with respect to visual input (in the X dimension), before (left) and after (right) training. X1, X2 and X3 code the negative, middle and positive range in the X dimension (threshold=0.3).

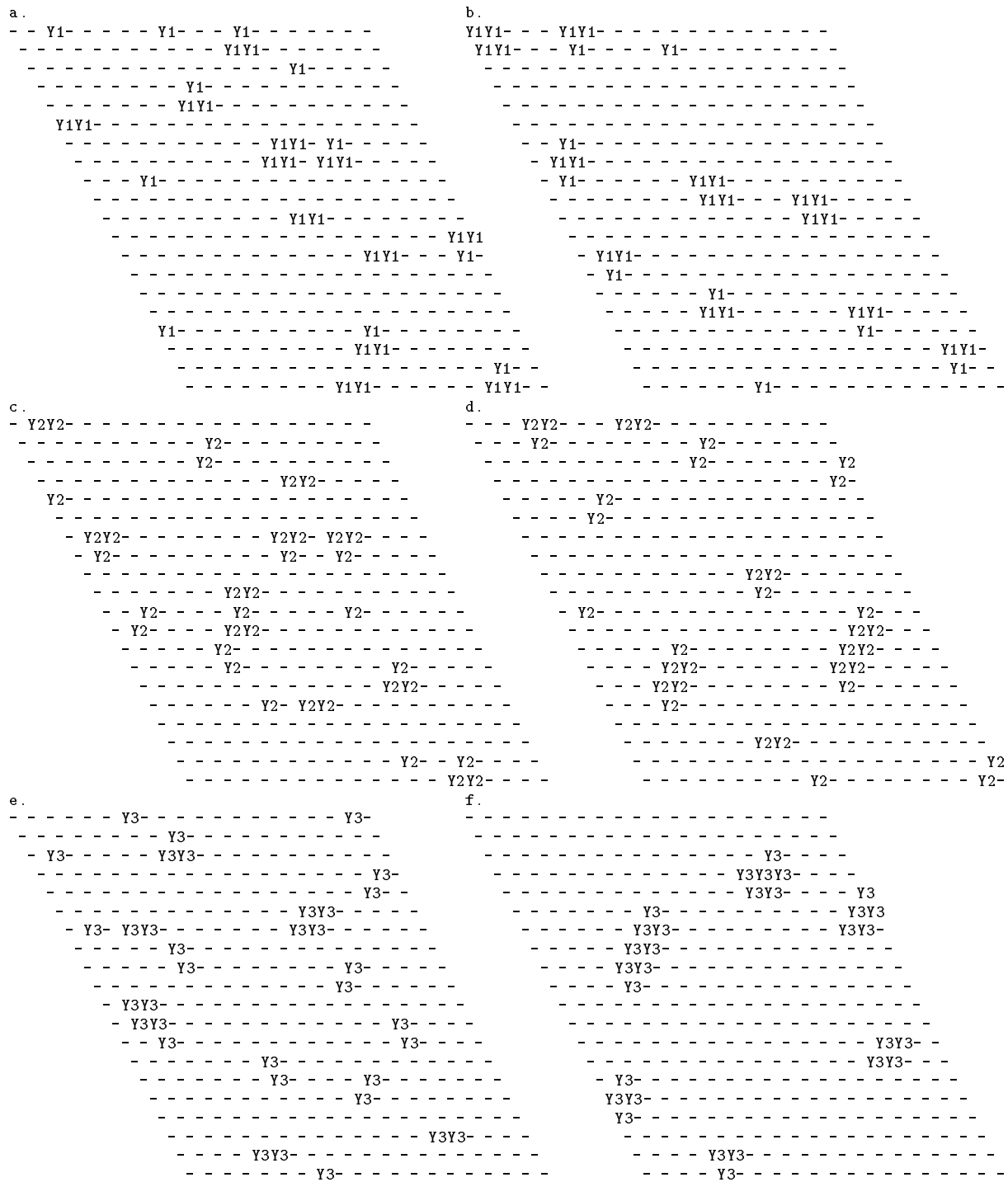


Figure A.20: The MI input maps with respect to visual input (in the Y dimension), before (left) and after (right) training. Y1, Y2 and Y3 code the negative, middle and positive range in the Y dimension (threshold=0.3).

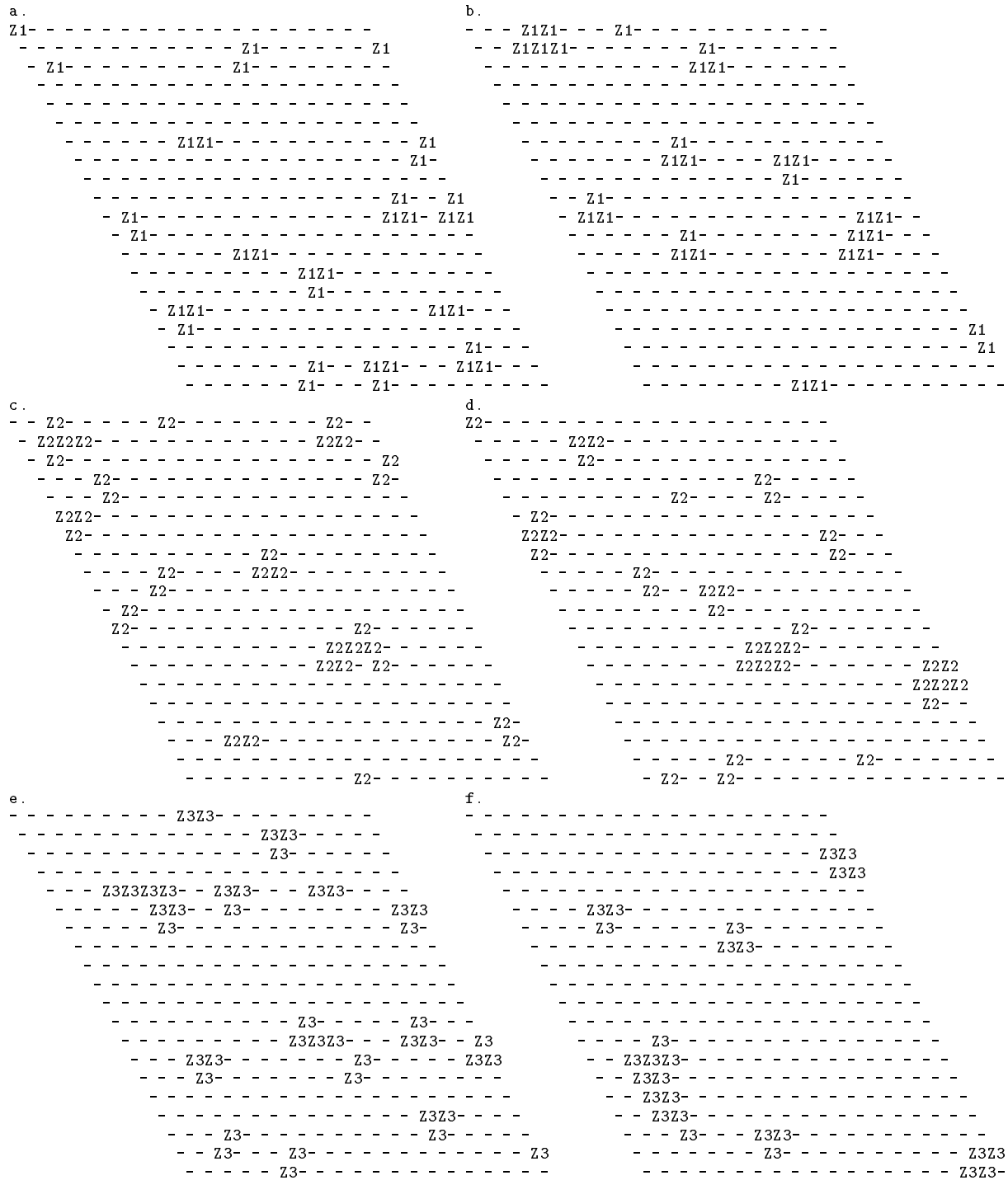


Figure A.21: The MI input maps with respect to visual input (in the Z dimension), before (left) and after (right) training. Z1, Z2 and Z3 code the negative, middle and positive range in the Z dimension (threshold=0.3).

### A.3 Cortical Feature Maps of Motor Control Model with Combined Proprioceptive Input and Visual Input

#### A.3.1 Proprioceptive input maps in PI layer

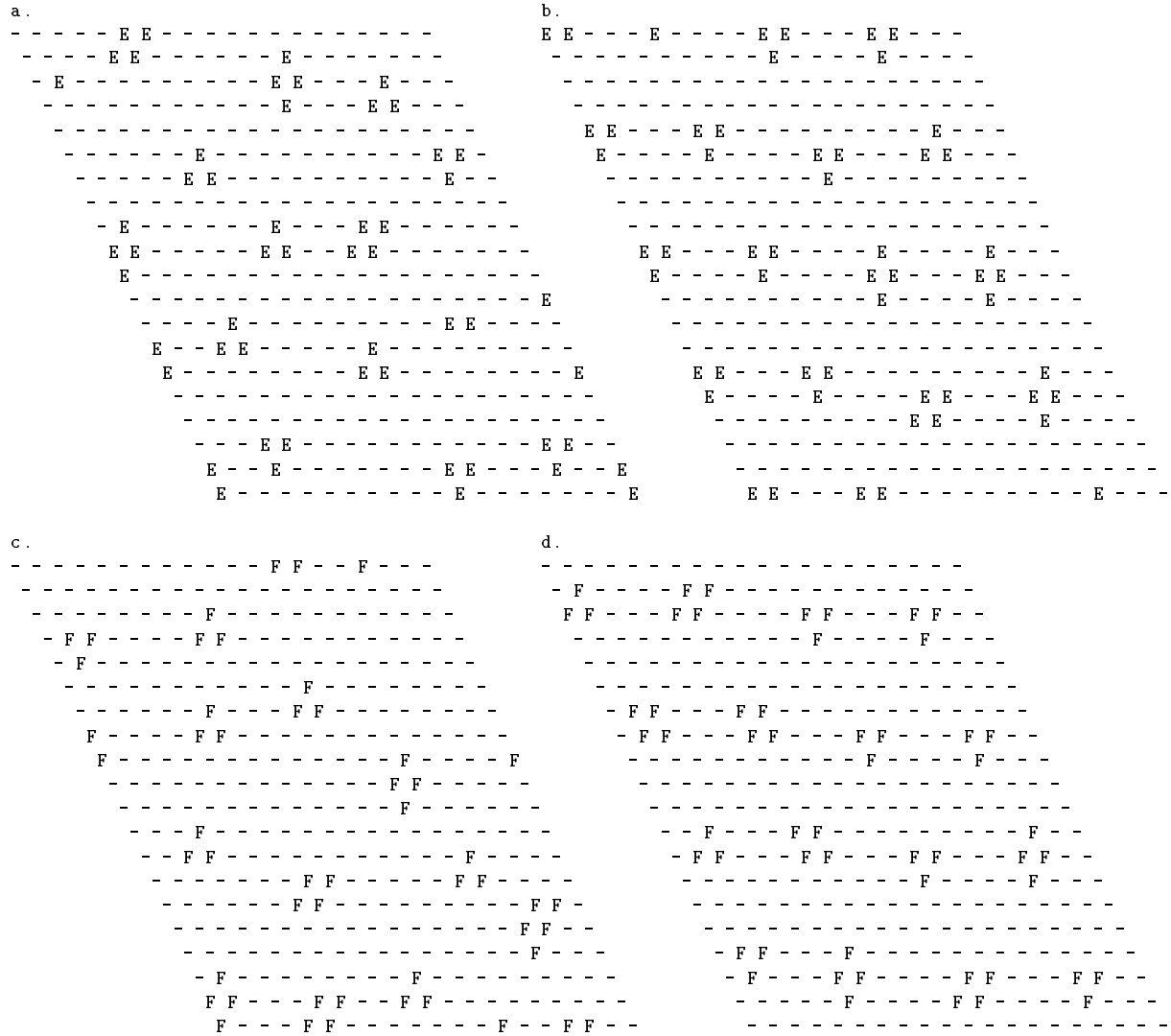


Figure A.22: Tuning of PI elements to the length of the upper arm extensor (E) and flexor (F) before (left) and after (right) training (threshold=0.2).

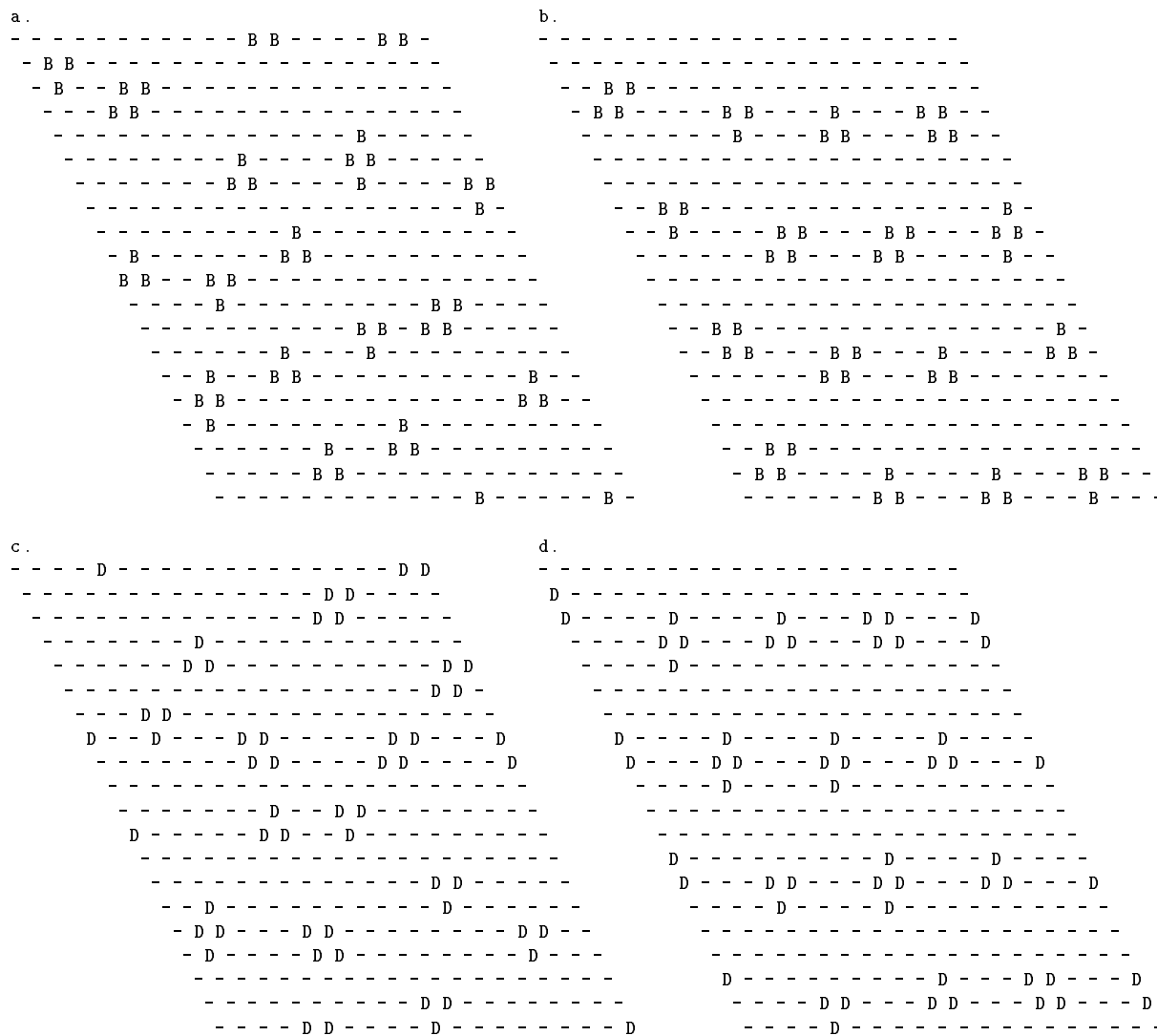


Figure A.23: Tuning of PI elements to the length of the upper arm abductor (B) and adductor (D) before (left) and after (right) training (threshold=0.2).



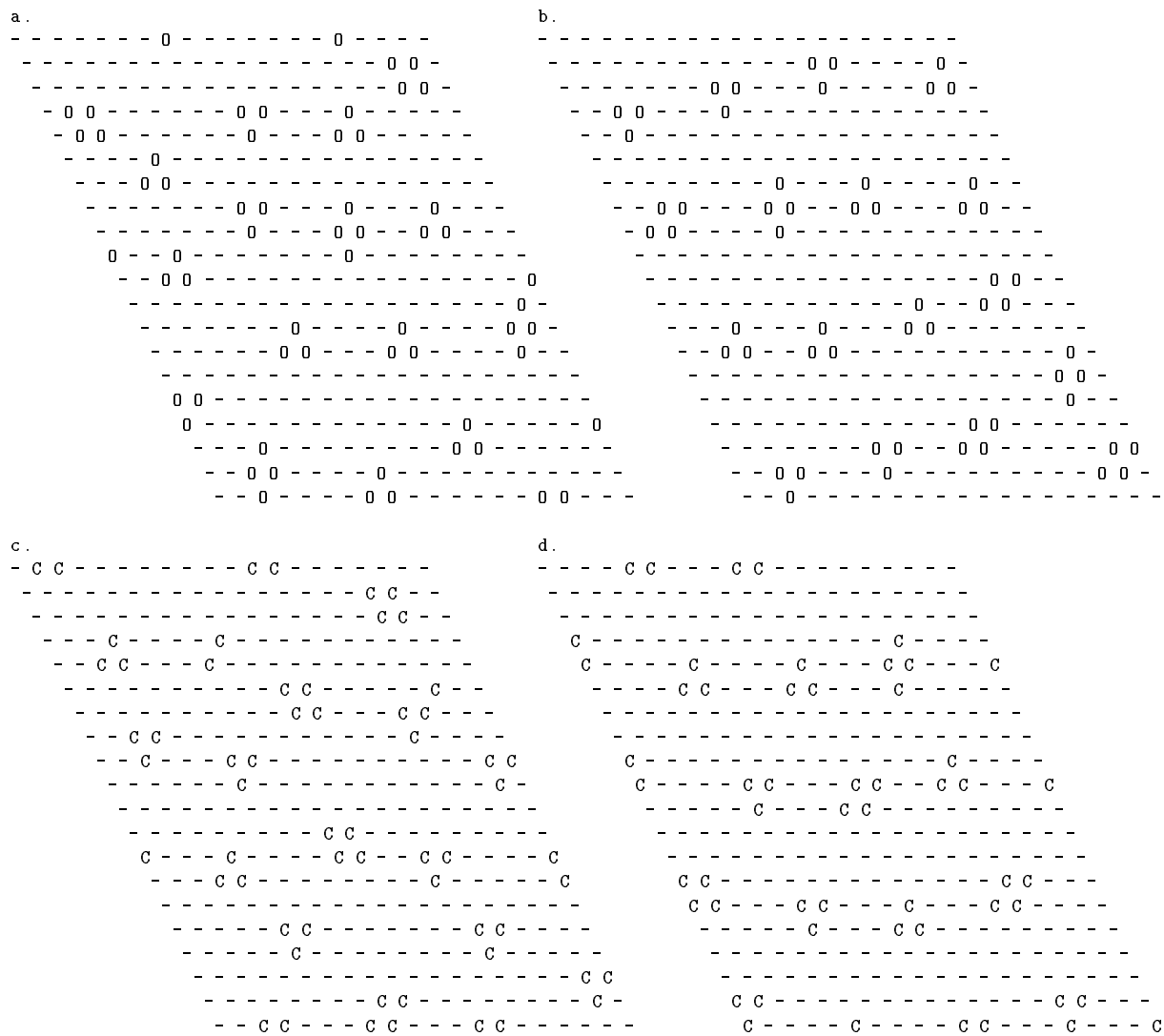


Figure A.24: Tuning of PI elements to the length of the lower arm extensor or operner (O) and flexor or closer (C) before (left) and after (right) training (threshold=0.2).

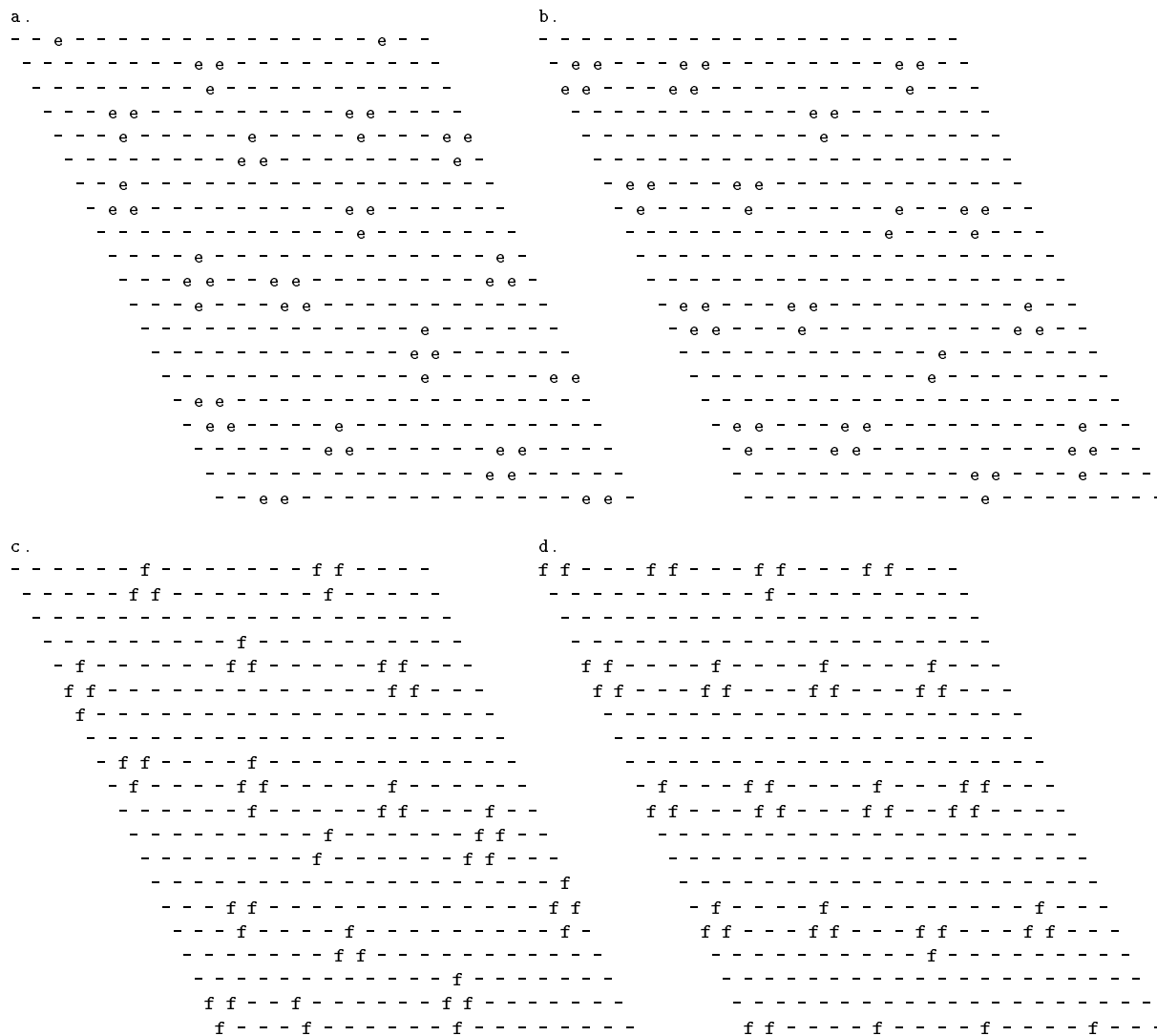


Figure A.25: Tuning of PI elements to the tension of the upper arm extensor (e) and flexor (f) before (left) and after (right) training (threshold=0.2).

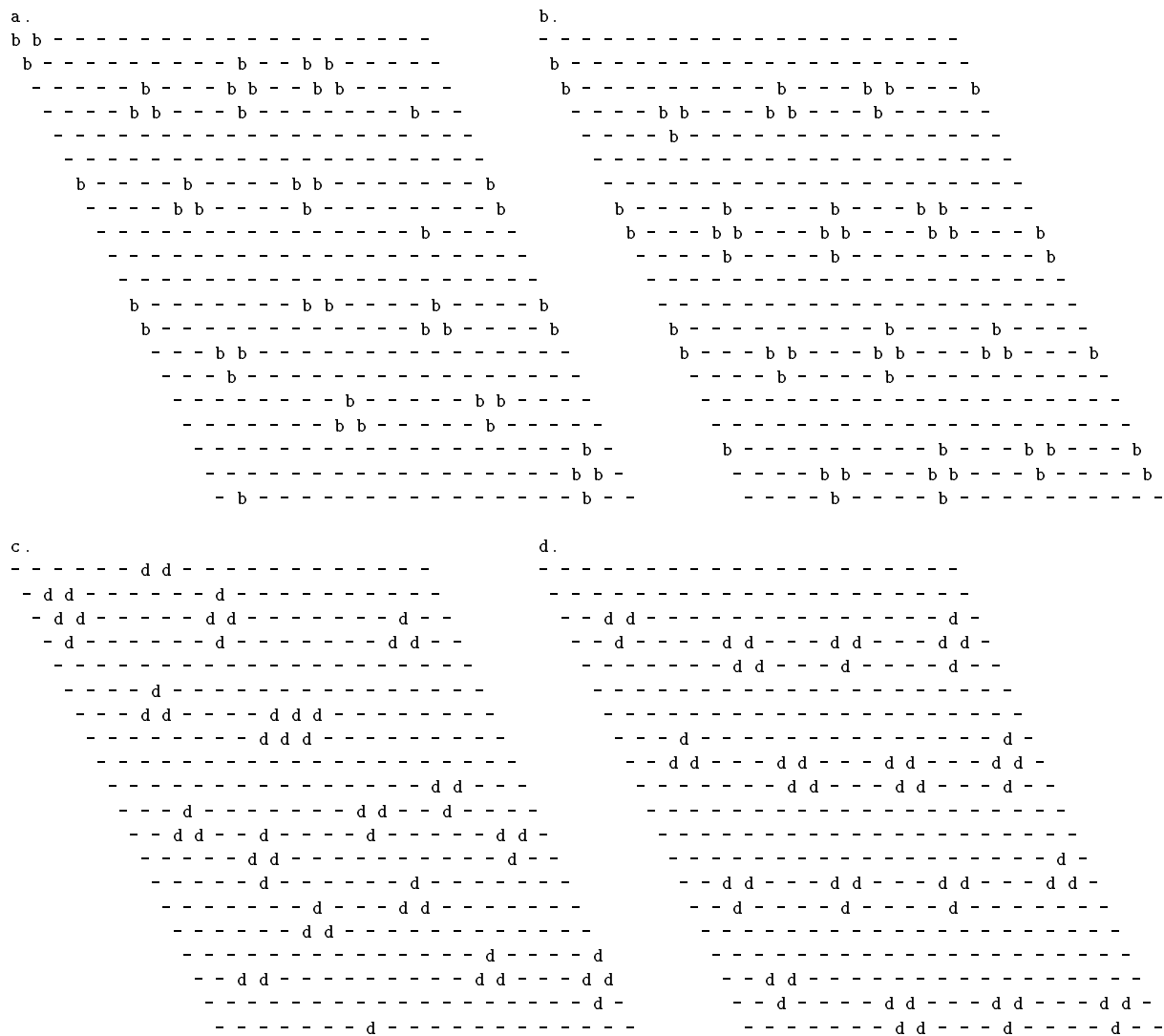


Figure A.26: Tuning of PI elements to the tension of the upper arm abductor (b) and adductor (d) before (left) and after (right) training (threshold=0.2).

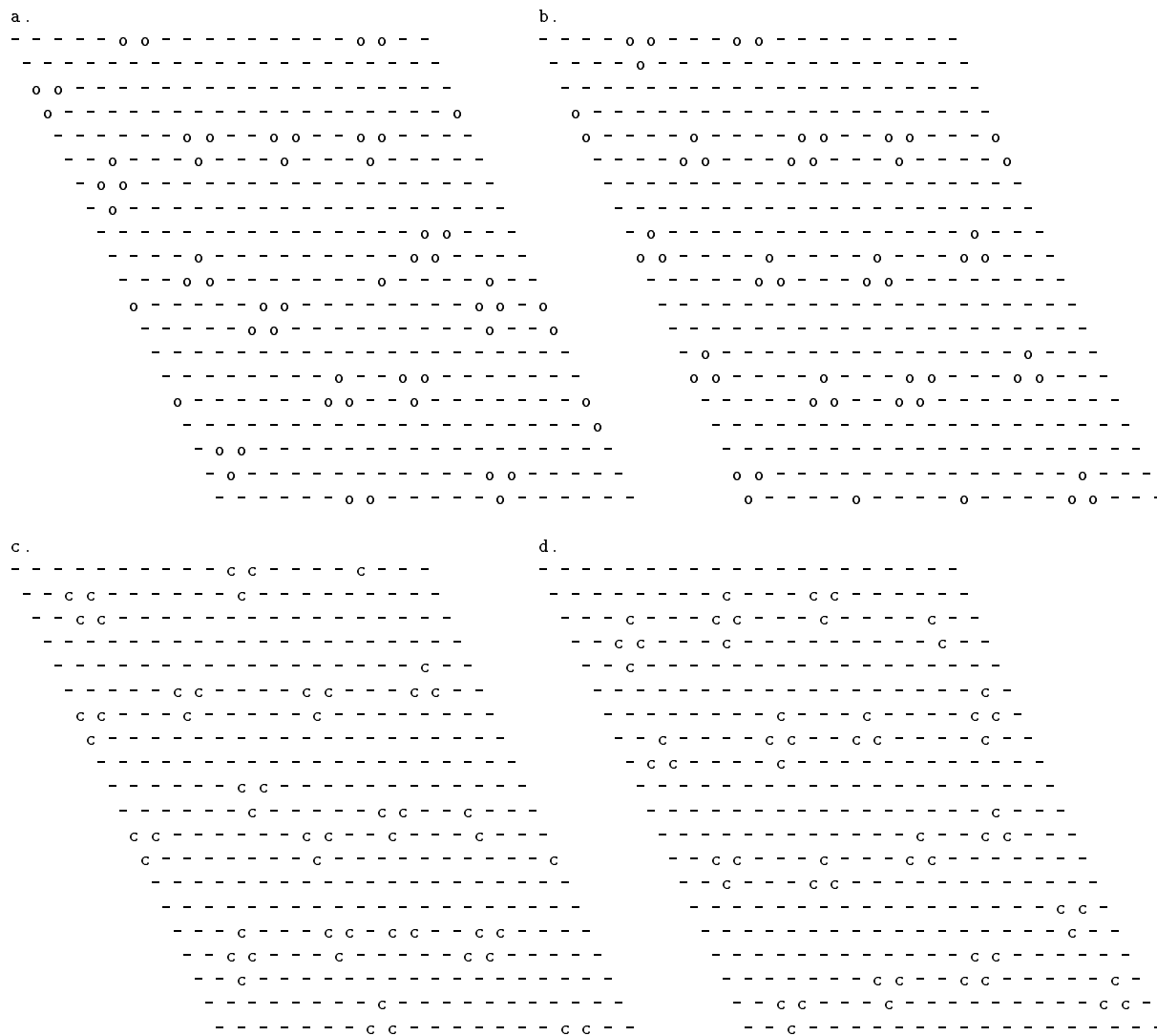


Figure A.27: Tuning of PI elements to the tension of the lower arm extensor or operner (o) and flexor or closer (c) before (left) and after (right) training (threshold=0.2).

### A.3.2 Proprioceptive input maps in MI layer

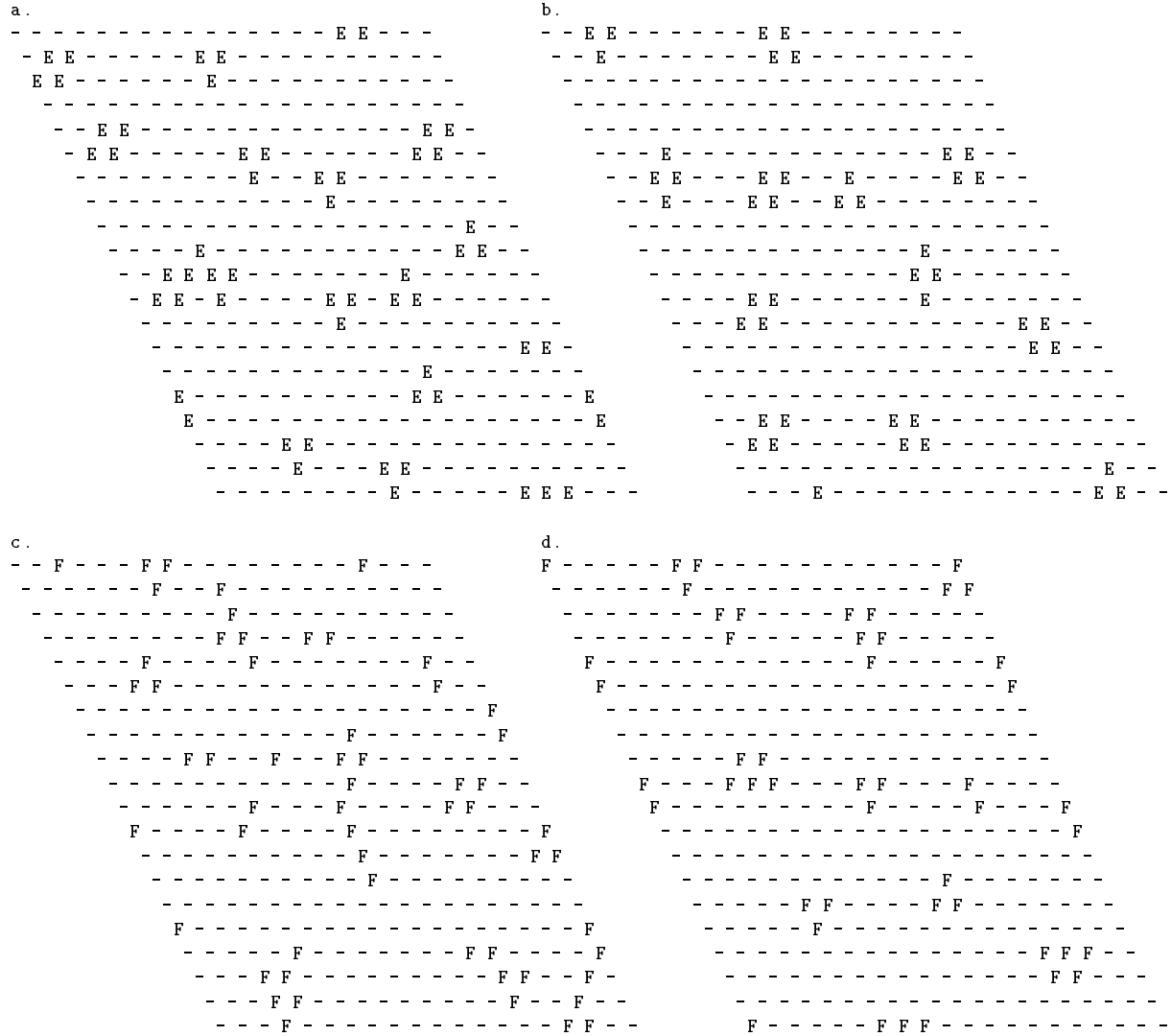


Figure A.28: Tuning of MI elements to the length of the upper arm extensor (E) and flexor (F) before (left) and after (right) training (threshold=0.4).

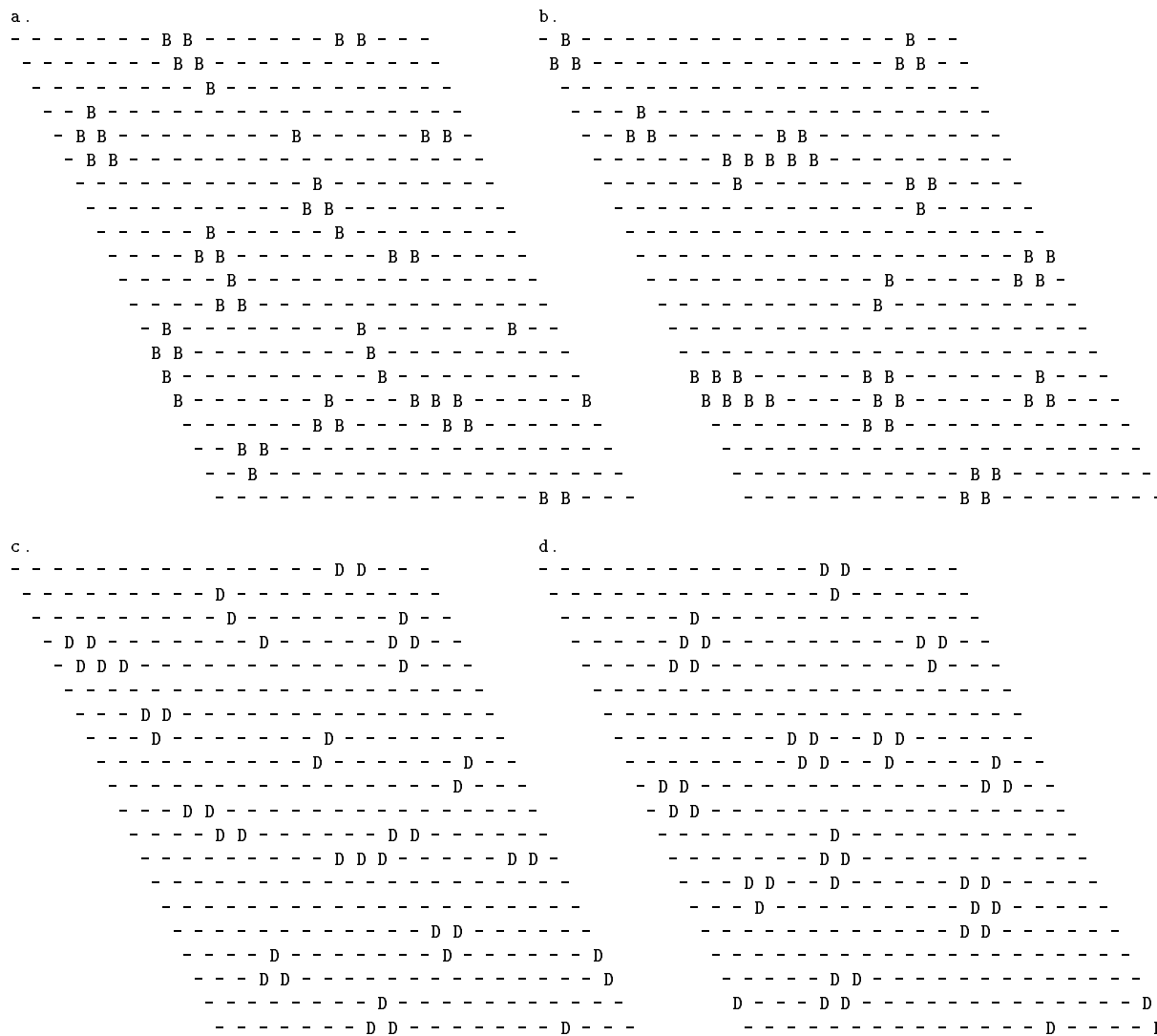


Figure A.29: Tuning of MI elements to the length of the upper arm abductor (B) and adductor (D) before (left) and after (right) training (threshold=0.4).

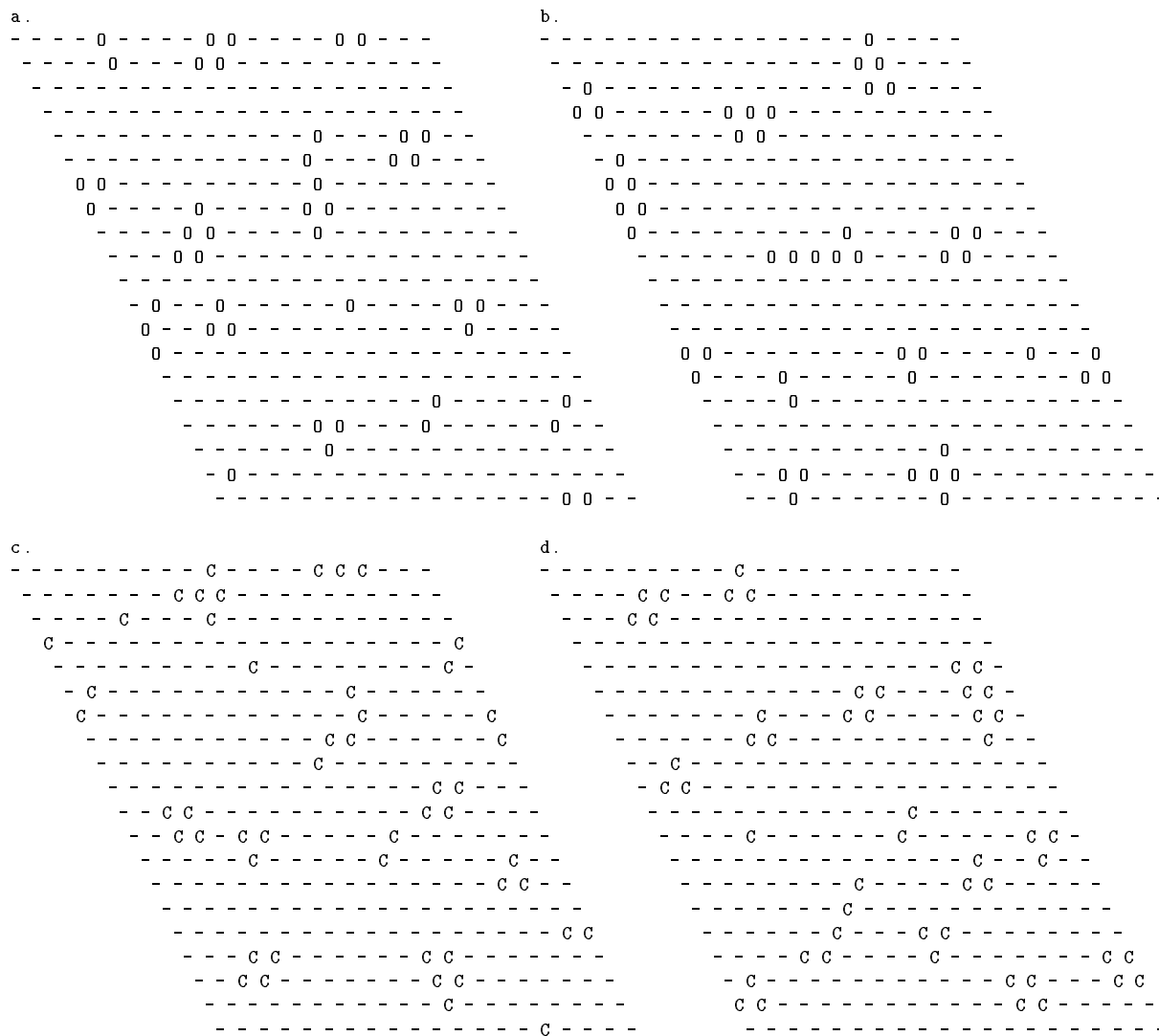


Figure A.30: Tuning of MI elements to the length of the lower arm extensor or operner (O) and flexor or closer (C) before (left) and after (right) training (threshold=0.4).

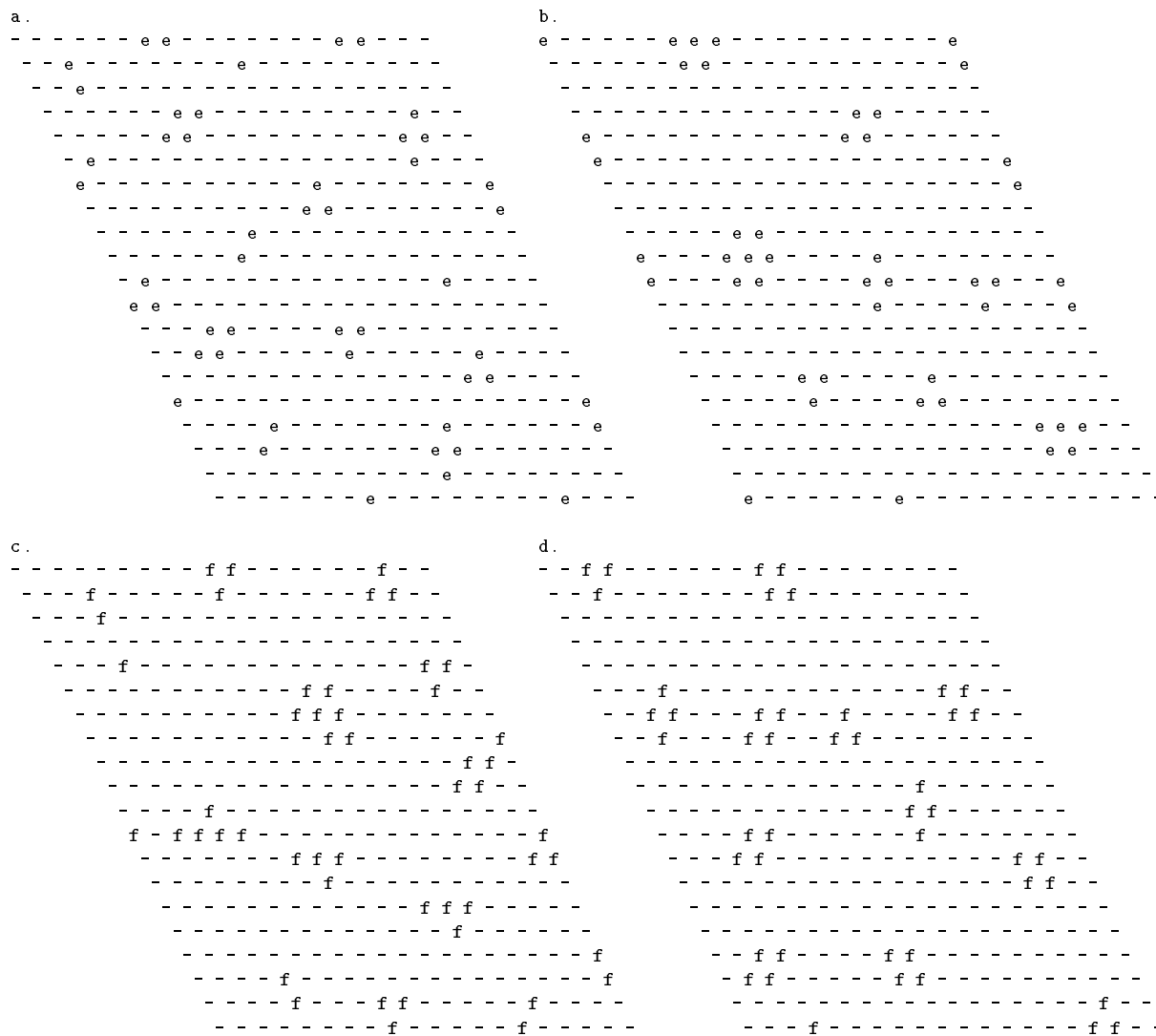


Figure A.31: Tuning of MI elements to the tension of the upper arm extensor (e) and flexor (f) before (left) and after (right) training (threshold=0.4).



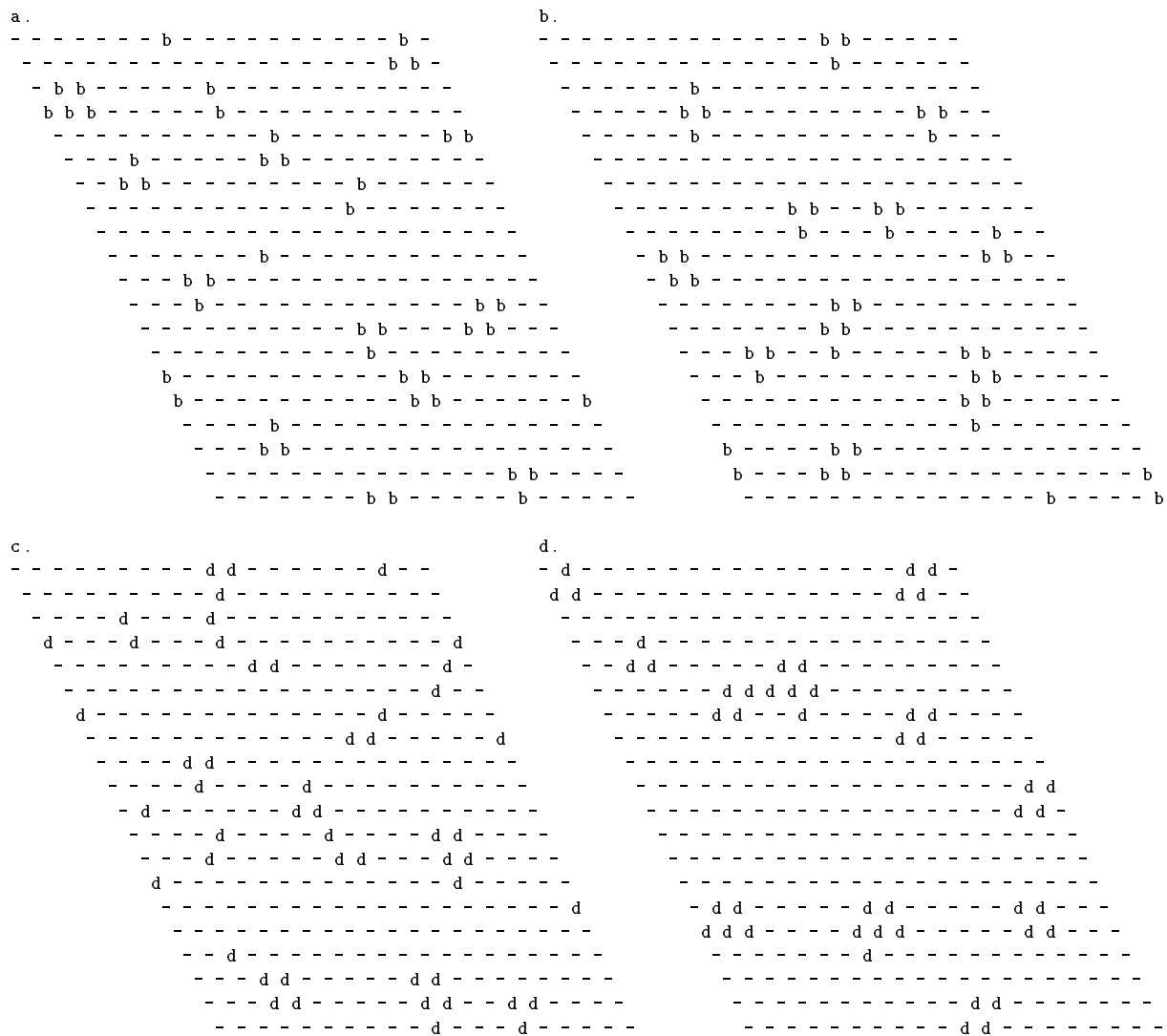


Figure A.32: Tuning of MI elements to the tension of the upper arm abductor (b) and adductor (d) before (left) and after (right) training (threshold=0.4).

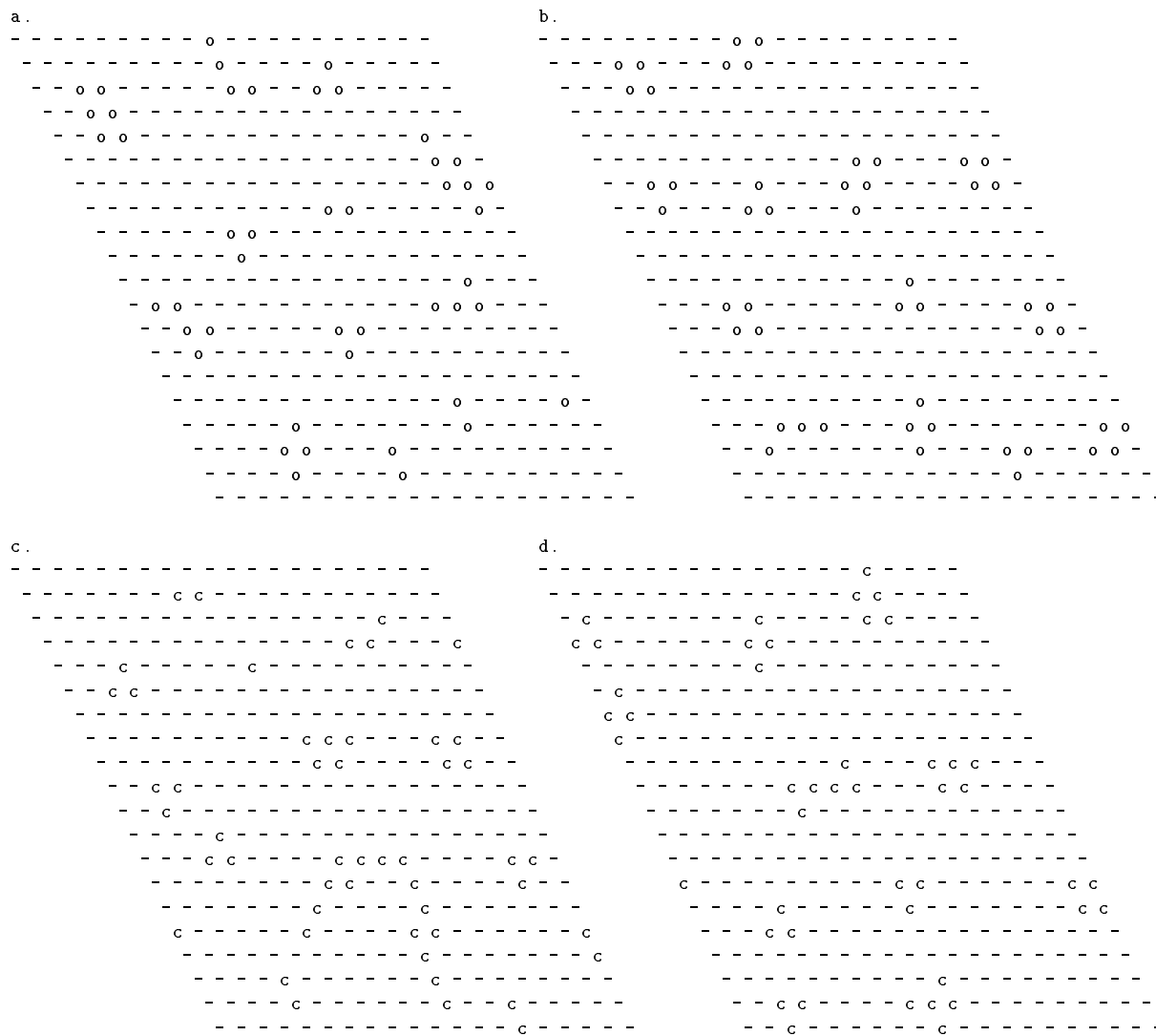


Figure A.33: Tuning of MI elements to the tension of the lower arm extensor or operner (o) and flexor or closer (c) before (left) and after (right) training (threshold=0.4).

### A.3.3 Motor output maps in MI layer

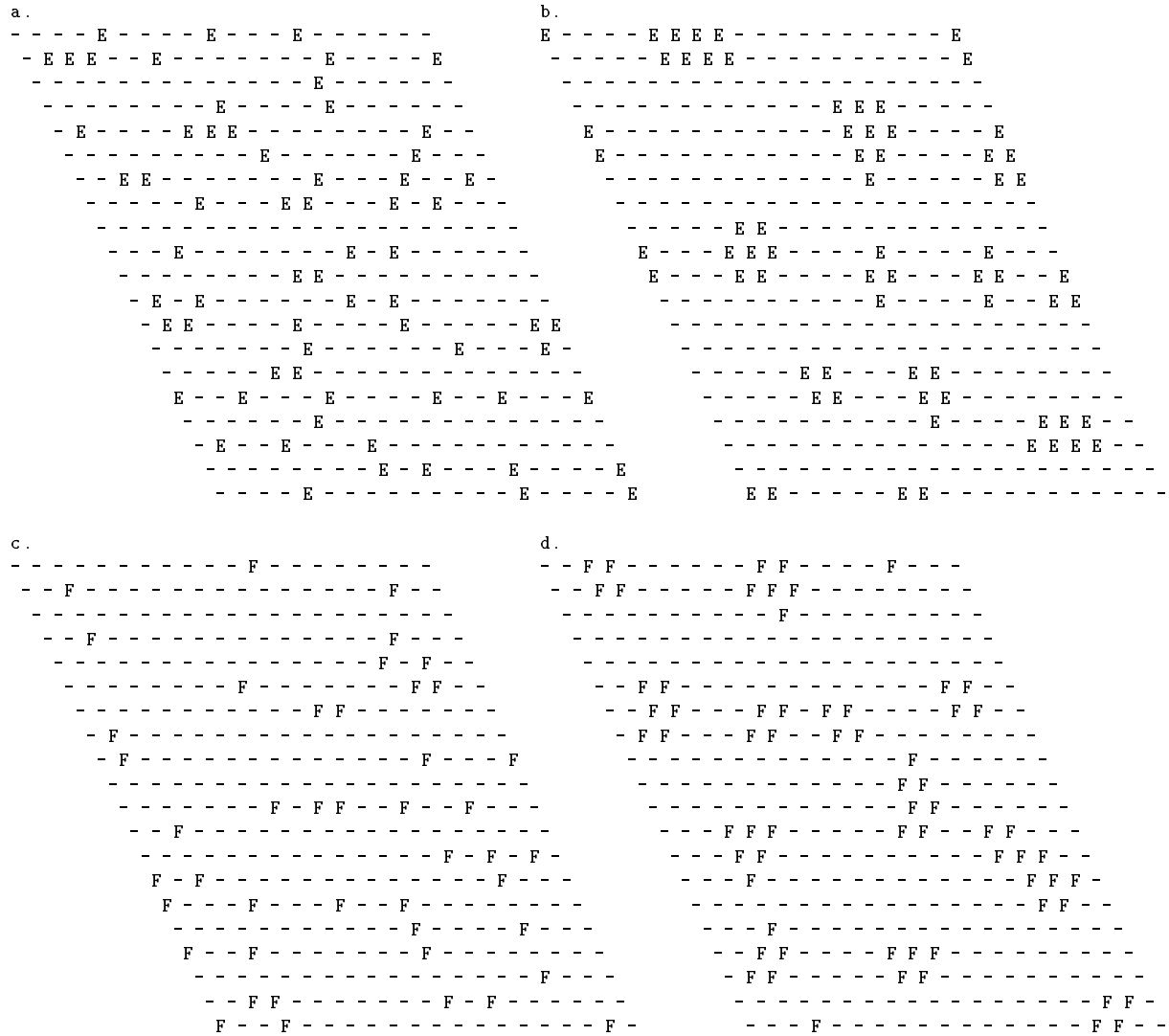


Figure A.34: MI output map before (left) and after (right) training for upper arm extensor (E) and flexor (F) (threshold=0.4).

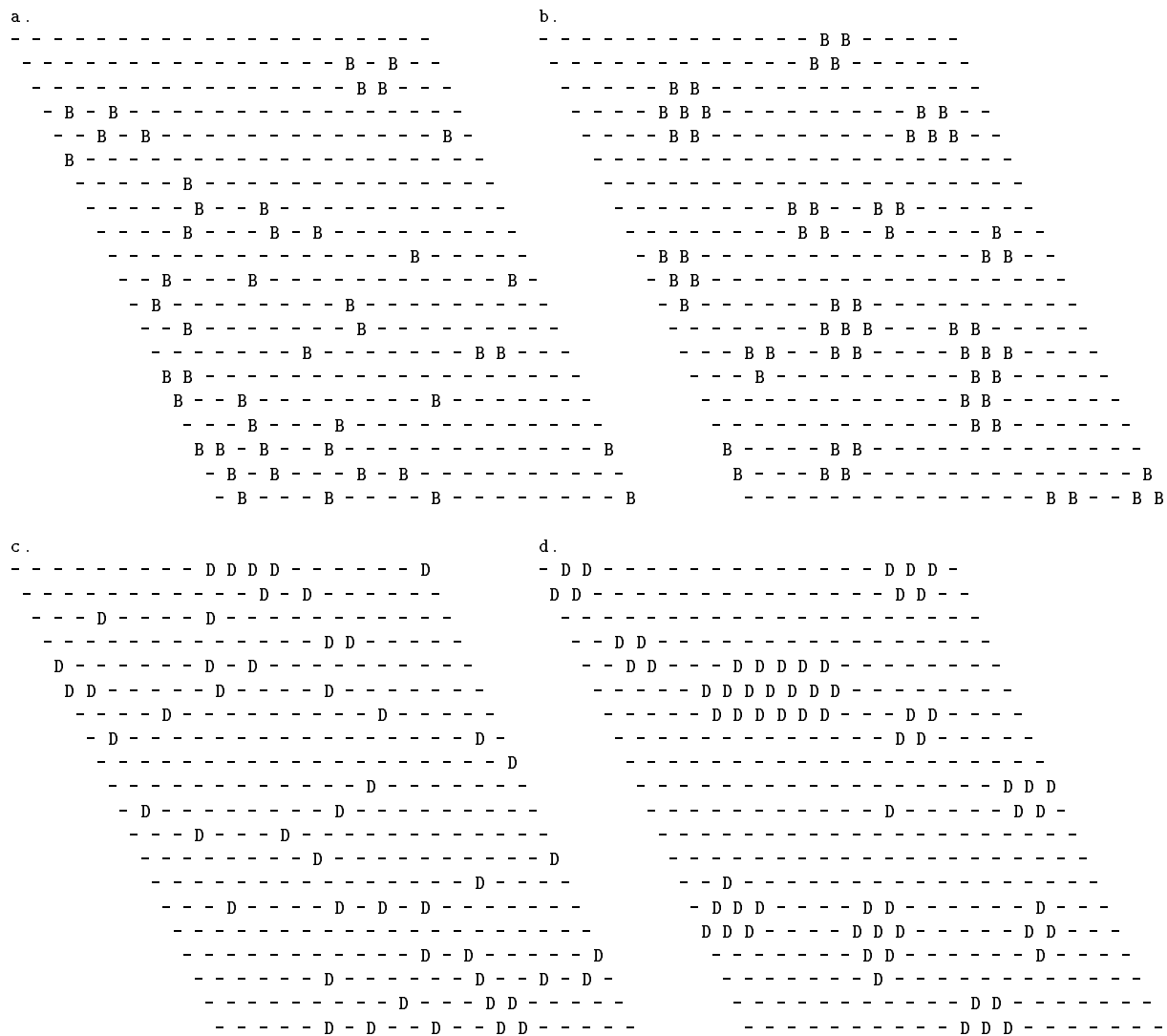


Figure A.35: MI output map before (left) and after (right) training for upper arm abductor (B) and adductor (D) (threshold=0.4).

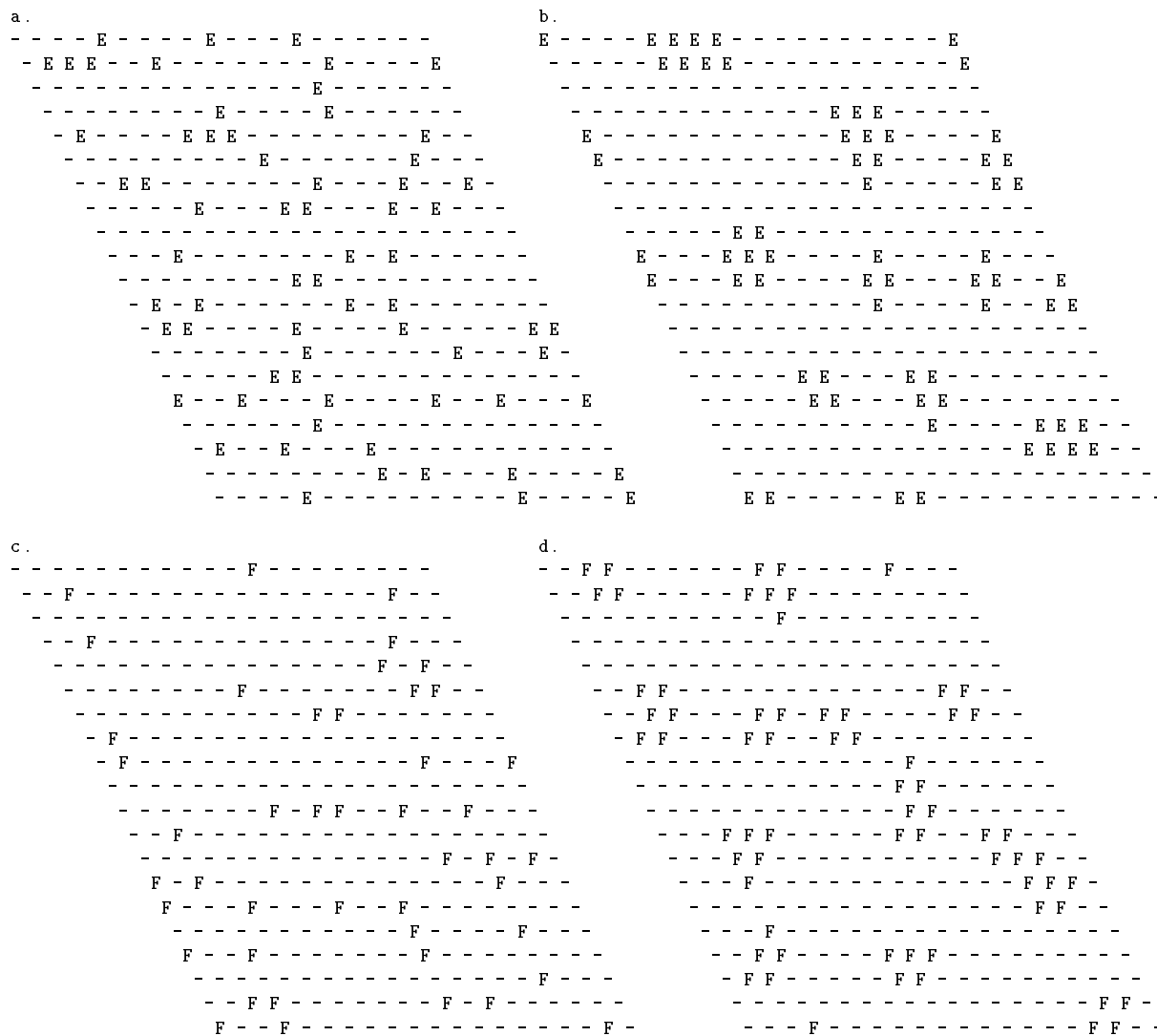


Figure A.36: MI output map before (left) and after (right) training for lower arm extensor or opener (O) and flexor or closer (C) (threshold=0.4).

### A.3.4 Visual input maps in MI layer

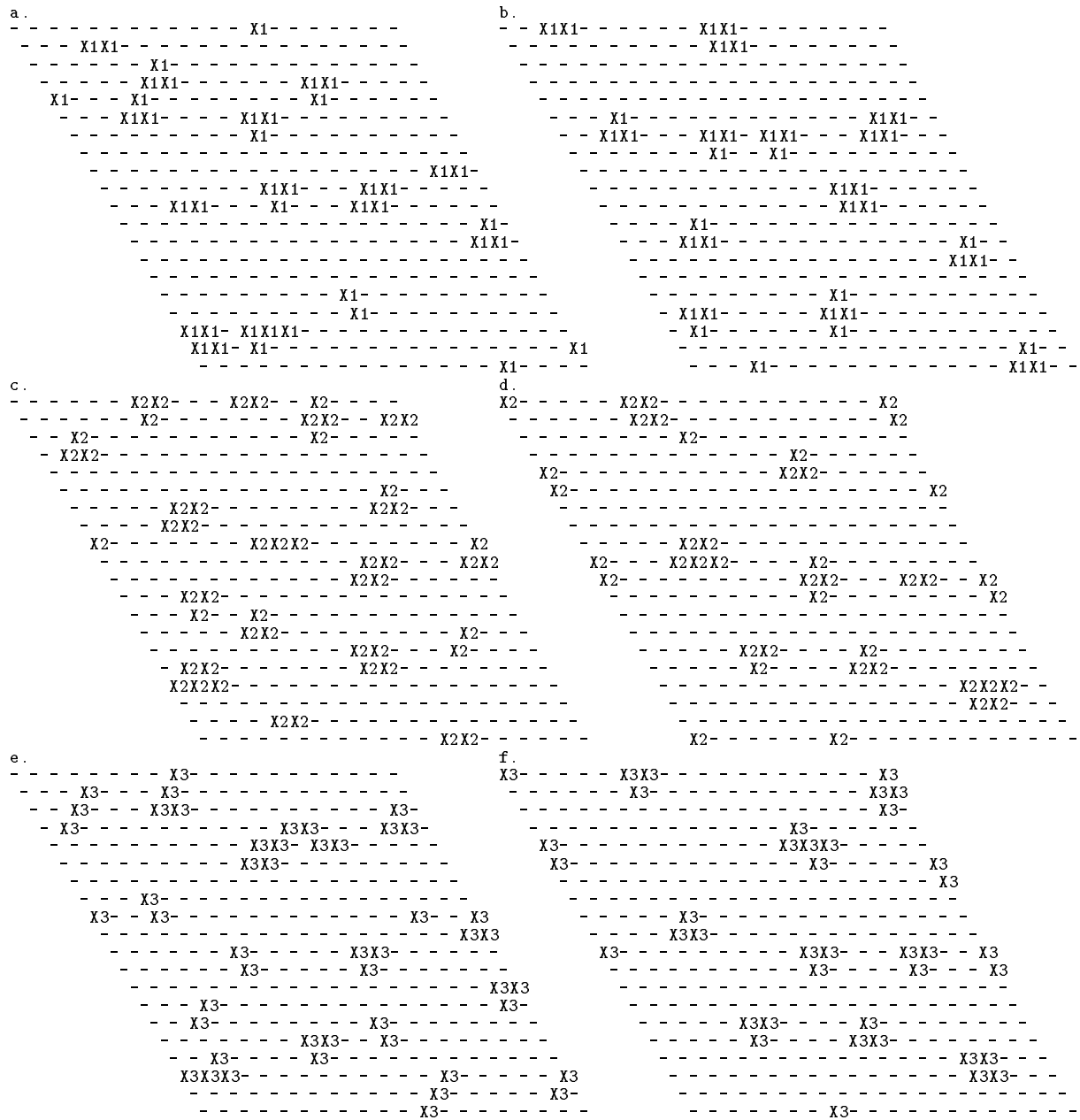


Figure A.37: The MI input maps with respect to visual input (in the X dimension), before (left) and after (right) training. X1, X2 and X3 code the negative, middle and positive range in the X dimension (threshold=0.3).

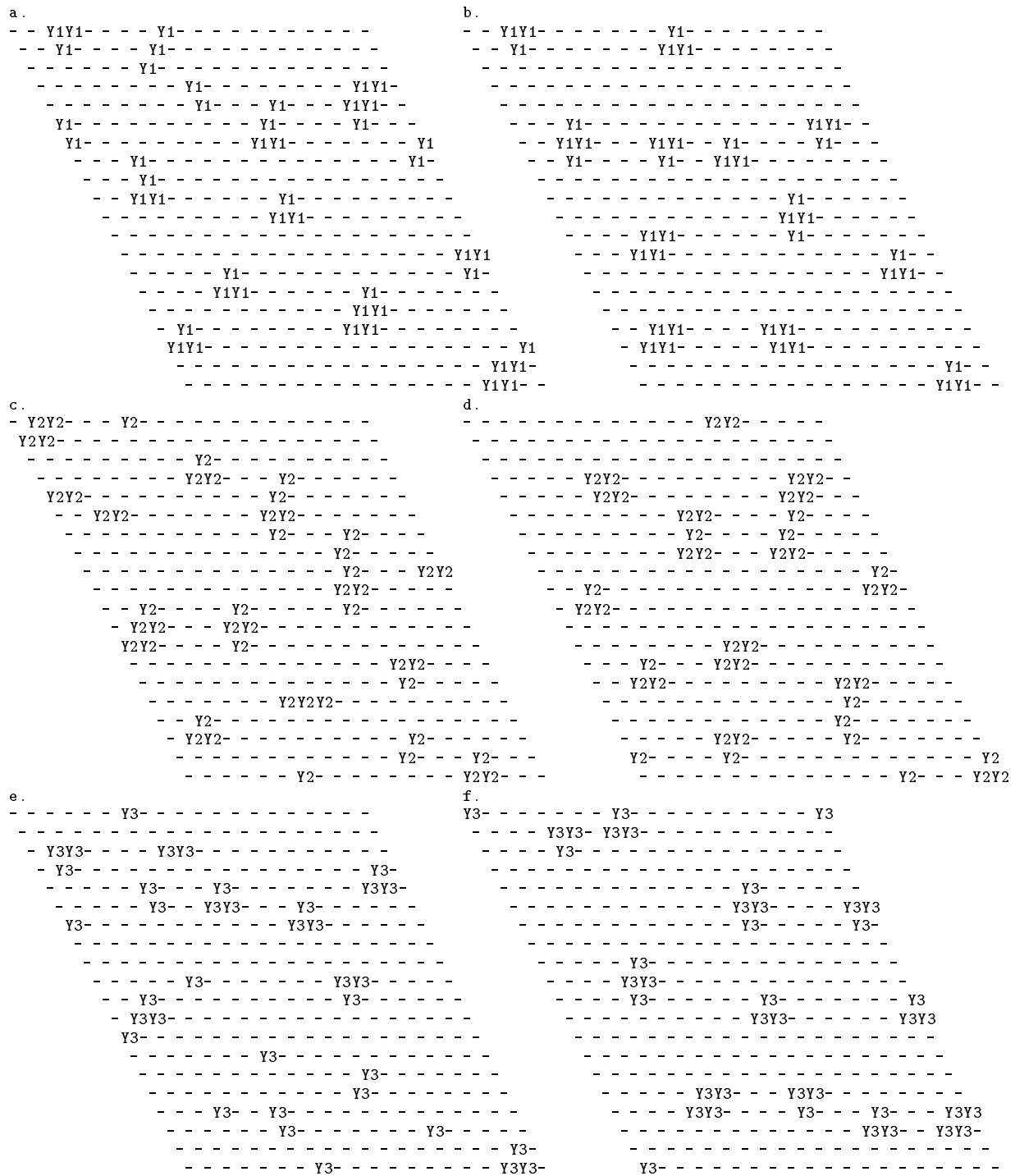


Figure A.38: The MI input maps with respect to visual input (in the Y dimension), before (left) and after (right) training. Y1, Y2 and Y3 code the negative, middle and positive range in the Y dimension (threshold=0.3).

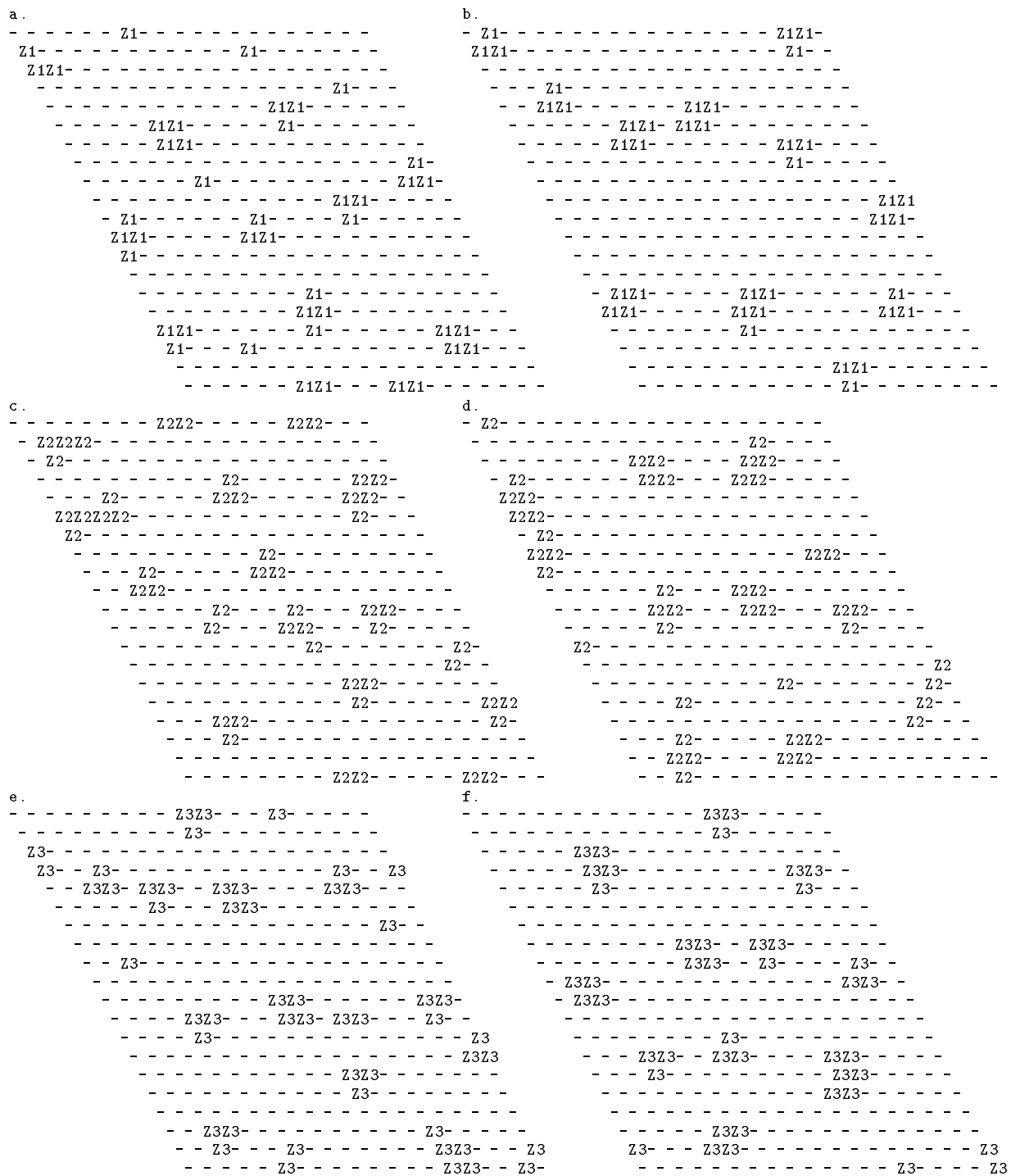


Figure A.39: The MI input maps with respect to visual input (in the Z dimension), before (left) and after (right) training. Z1, Z2 and Z3 code the negative, middle and positive range in the Z dimension (threshold=0.3).



## Appendix B

### Lesioning the Motor Control Model

In this appendix I briefly summarize the results of some simulations examining the effects of sudden, focal lesions of varying sizes to the cortex regions of the motor control model (“simulated ischemic strokes”). This summary is included here to illustrate an application of the model developed in this dissertation work, specifically to study questions about how the brain might recover following an ischemic stroke. The work described in this appendix was done in collaboration with my colleagues in our research group. A more complete description can be found in [Goodall *et al.*, 1997].

Two sets of simulations were done in which an area of focal damage was suddenly imposed upon a previously trained network. The first set involved lesions in PI, the second lesions in MI. In both cases, a focal lesion was simulated by clamping the activation levels of a contiguous set of “lesioned” cortical elements permanently at zero. In addition, connections to and from lesioned cortical elements were severed.

The effect of each lesion on the existing proprioceptive and motor maps in the trained, intact cortex was examined twice: immediately after the lesion, and after continually training the network with 2000 additional random input stimuli in MI. An analysis of changes in the position of the model arm following cortical stimuli was also made both immediately post-lesion and after further training. All lesion effects were compared with the pre-lesion network as well as with a control network. The control network was an exact copy of the intact pre-lesion network made immediately before lesioning. Training was continued with this unlesioned control model, with additional random input stimuli, so that any map alterations due to continued training alone could be compared to those due to lesioning plus continued training.

#### B.1 Unlesioned Model

Fig. B.1a shows the model arm in four of six test positions, for the intact pre-lesion model, corresponding to “requests” to contract the upper arm extensor, upper arm flexor, upper arm abductor and upper arm adductor. As seen in Fig. B.1a, the four arm positions corresponding to these motor cortex stimuli are in the anticipated directions and are virtually indistinguishable for the trained prelesion model (dotted lines) and the further trained control model (hatched lines). The stability of both cortical maps and arm positioning in response to cortical stimuli in the control model indicate that changes seen in the lesioning simulations described below are caused by the lesions themselves.

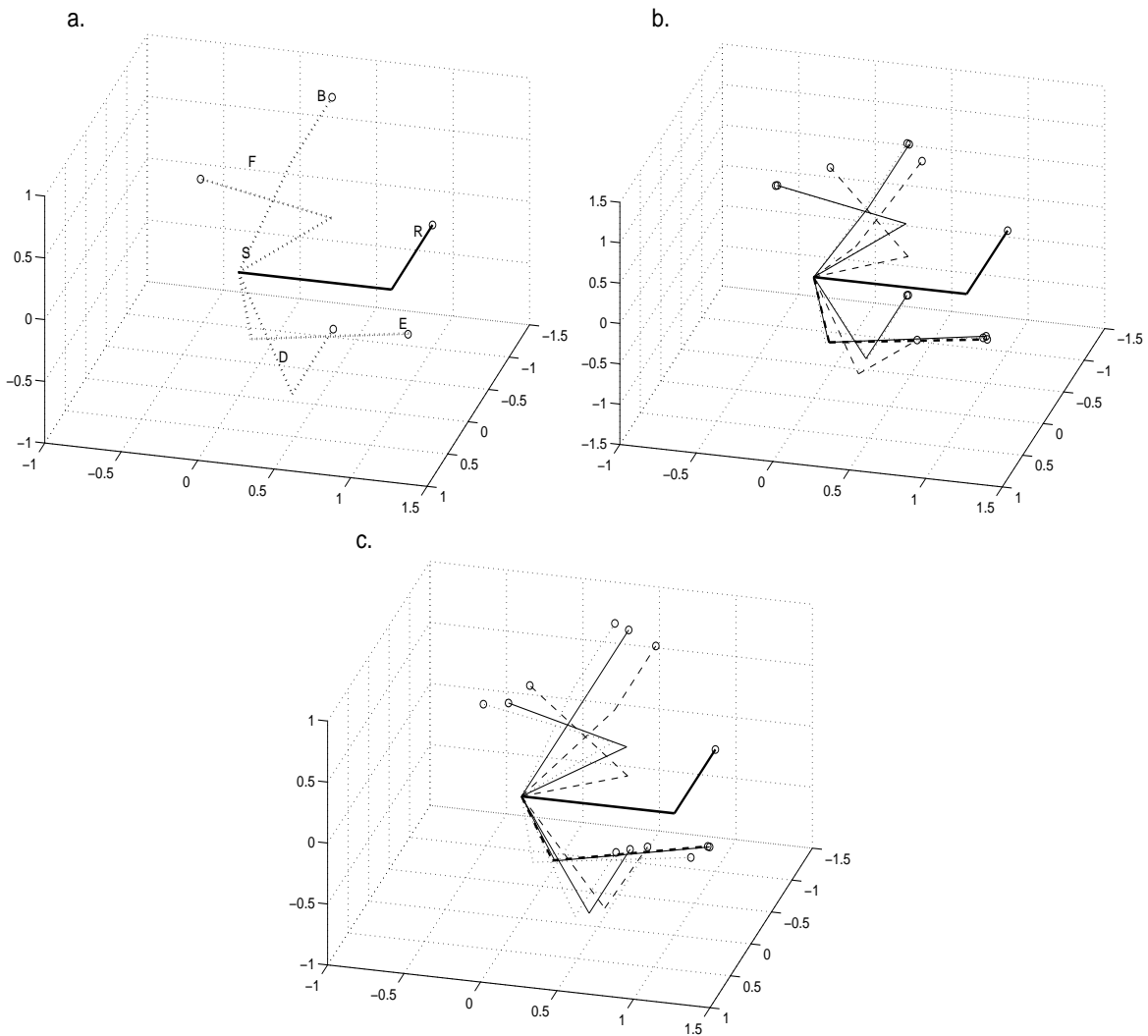


Figure B.1: (a) Position assumed by the model arm at rest in the absence of external stimuli (R; thick solid line) and in response to four of six cortical test stimuli. The arm is on the right side of the body and is viewed from the back (S = right shoulder; small circles = hand positions). Each test stimulus provides external input to cortical elements in MI that are most strongly connected to a specific muscle group (here upper arm extensor E, flexor F, abductor B and adductor D). For example, activating upper arm abductor elements in MI elevates the arm to position B. The positions assumed by the arm in response to cortical stimuli are appropriate and indistinguishable for the trained prelesion (dotted lines) and control (cross hatching) states of the model. Similar results are found for the lower arm flexor and extensor (not shown). (b) Arm positions for an 8x8 focal lesion of PI shown pre-lesion (dotted line; largely obscured by overlapping solid lines), immediately post-lesion (dashed line) and after 2000 further random input stimuli in MI (solid line). (c) Arm positions for a 16x16 lesion of MI, pre-lesion (dotted line), immediately post-lesion (dashed line), and after 2000 further random input stimuli in motor cortex layer MI (solid line).

## B.2 Focal Lesions in Proprioceptive Cortex

The effects of structural lesions in PI were examined under a variety of conditions. Changes to the feature maps in PI were observable immediately after a structural lesion occurred in this layer, as the first phase of a two-phase reorganization process. Following the primary structural lesion in PI, the activity of surrounding elements was decreased, forming a secondary functional lesion. For example, Fig. B.2a shows a perilesion zone of relatively inactive cortical elements (marked by “-”s) seen immediately following an 8x8 focal lesion; these elements do not respond to the stretch of any of the muscles above a threshold of 0.4.

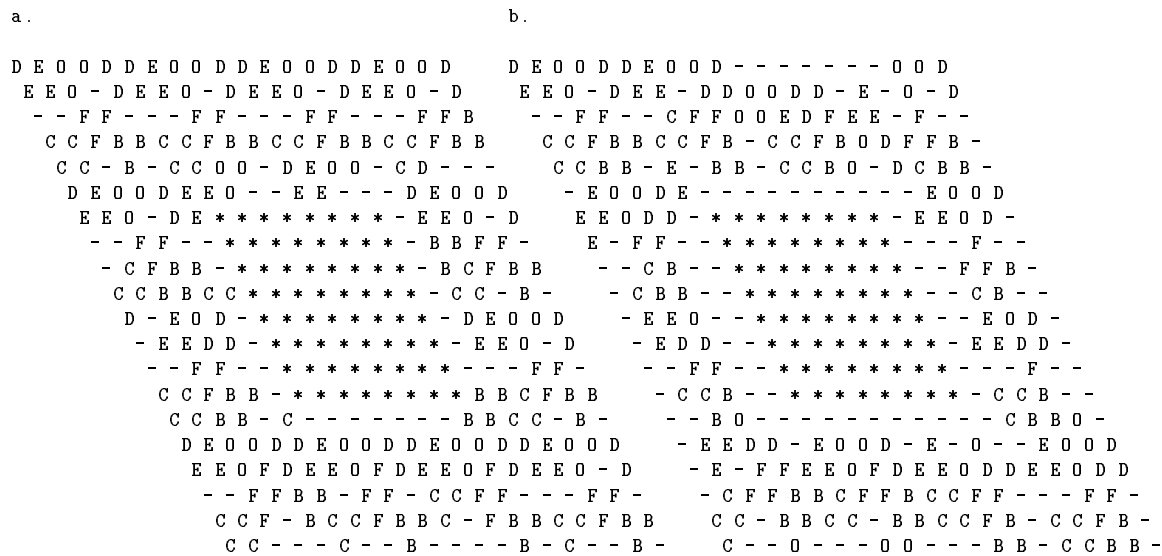


Figure B.2: Muscle stretch map of proprioceptive cortex layer PI (above threshold 0.4) for an 8x8 focal lesion of PI (a) immediately post-lesion, and (b) following 2000 further random input stimuli in motor cortex layer MI. Asterisks indicate the imposed structural lesion, adjacent “-”s the functional perilesion deficit. A partial but less pronounced “ring” of poorly responsive units is evident at a distance 6 from the lesion (outer border of (b)). This is not an “edge effect”; its genuine presence was verified with a larger cortical region.

The second phase of reorganization occurred more slowly with continued synaptic changes during the post-lesion period. With time, as the map reorganized in the context of continued proprioceptive input and synaptic changes, the functional lesion gradually enlarged. For example, with an 8x8 structural lesion there was a 77% increase in perilesion inactivity at distances 1 and 2 from the lesion edge over the long term (see Fig. B.2b, in comparison with Fig. B.2a). Similar changes were observed with the proprioceptive map of muscle tension. Over time, clusters of elements responsive to the stretch of a particular muscle also shifted position in the feature map.

The functional lesion effects described above occurred largely independently of structural lesion size in PI. They are representative of the effects observed with lesions that varied incrementally in size from 2x2 to 8x8. The dynamics of these functional lesions can be analyzed further by examining the mean activation level of cortical elements, averaged over all of the test input patterns. There was

an essentially uniform pre-lesion mean activation of the PI elements of roughly 0.12. Immediately following the structural lesion, the mean activation level of cortical elements directly adjacent to the lesion site dropped to 0.08, about 70% of its pre-lesion value. With additional synaptic modifications following the lesion, these perilesion effects in the PI layer were intensified (about 25% of prelesion value) and shifted outwards.

Further examination of the model, following lesions in PI, reveals that perilesion cortical elements were activated essentially the same amount for all input stimuli, in contrast with the pre-lesion cortex where elements were activated selectively for some specific input stimuli but not others. This uniformity occurred as the result of the loss of excitatory support from cortical elements in the structural lesion via intracortical connections. As the map reorganized following the lesion, the weights to these perilesion cortical elements tended to become uniform.

Immediately following the larger structural lesions (5x5 and larger) in PI, an irregularly shaped area of inactive motor cortex elements appeared in the center of the sensory maps of the MI layer, and did not resolve with further training. Given the coarsely topographic projections from PI to MI (projections from PI to MI elements within a radius 4), the observed inactive zone in the center of the motor cortex sensory map is expected, and can be viewed as an example of diaschisis. In addition to these effects on the sensory maps of the MI layer, larger PI lesions produced a central region in the motor output map that did not activate any muscle groups in the lower motor neuron layer. This was due to the loss of excitatory input to this region from the corresponding lesioned area in PI. The percentage of MI elements activating one or more muscle group(s) in the motor output maps was 77% prior to lesioning. This decreased with larger PI lesions (5x5 and larger), e.g., with an 8x8 PI lesion, the percentage dropped to 68% over time.

The decrease in motor output map responsiveness with lesions of increasing size led to “weakness” of the model arm following a lesion in PI. Fig. B.1b shows the arm position for the same four test inputs to MI as in Fig. B.1a, for an 8x8 focal lesion in PI. Immediately post-lesion, a measurable shift was observed in arm positions away from their pre-lesion position and towards the neutral, resting position of the arm. For example, the elbow position immediately post-lesion for the upper arm flexor test was  $20^\circ$  away from its pre-lesion position, revealing a weakened flexor response. Similar weakened responses were seen with the contraction tests of the abductor, adductor and lower arm flexor immediately post-lesion. This occurred due to functional loss of MI elements that activated each muscle group. However, over time with continued cortical plasticity, the arm positions for all test inputs realigned with their pre-lesion positions, representing essentially complete “recovery”. With larger PI lesions, e.g., 16 x 16, such recovery was incomplete.

### **Focal Lesions in Motor Cortex**

A separate set of simulations was performed to study reorganization of the MI cortical maps following focal structural lesions of varying sizes in MI (2x2 to 8x8). For sufficiently large lesions, reorganization after a structural lesion in MI was seen in both the MI sensory and motor output maps. Immediately after such large focal lesions to MI, both the stretch and tension sensory maps for MI adjusted so that there was an increase in the number of responsive elements in normal cortex near the lesion edge. In contrast to PI lesions, no perilesion zone of decreased activation was present, as can be seen in Fig. B.3a. At distances 1 and 2 from the lesion edge there was an increase in the number of responsive elements over pre-lesion levels, from 91% pre-lesion to 96% immediately after this 8x8 lesion. Although the change in absolute numbers of responsive elements is small, it accurately reflects a substantial increase in mean activation levels of all elements averaged over all

inputs in this perilesion zone (from 0.14 before lesion to 0.21 after). Over time, the distance 1 and 2 responsiveness stabilized at 99%, as is seen in Fig. B.3b. Overall rates of responsiveness for the MI sensory maps increased slightly immediately following the onset of the lesion, but then dropped back to prelesion levels with continued post-lesion synaptic modifications.

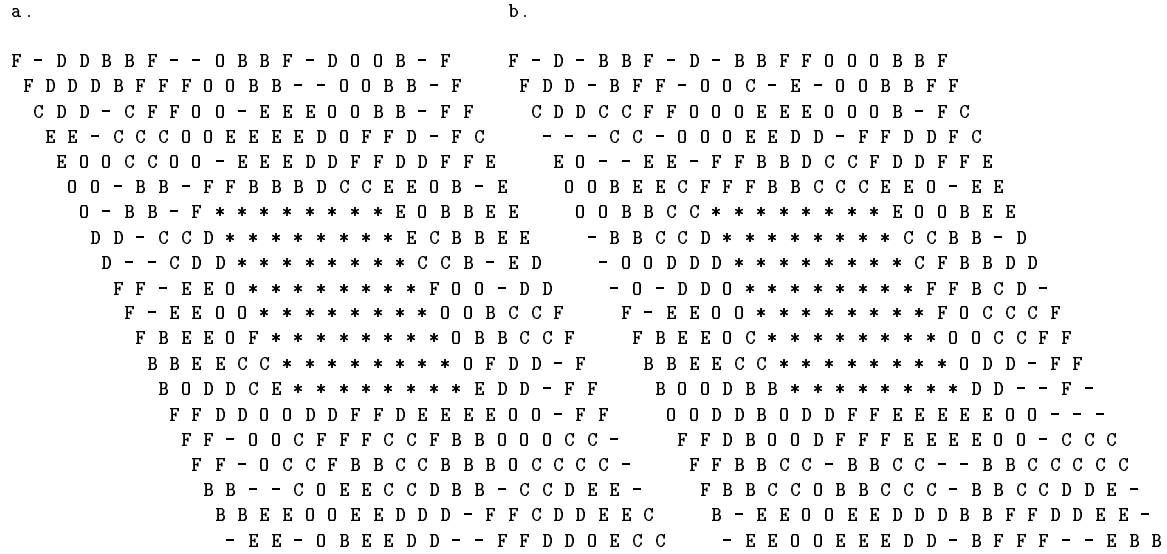


Figure B.3: Muscle stretch map of motor cortex layer following an 8x8 lesion in MI (a) immediately post-lesion, and (b) after 2000 further input stimuli in MI.

This post-lesion reorganization result is similar to results of prior studies of structural lesions to cortical layers with topographically-ordered somatosensory inputs [Armentrout *et al.*, 1994]. In this context, it is important to note that the topographically-ordered connections between PI and MI in this current model are similar to those between thalamus and sensory cortex in the earlier model (projections from PI to corresponding MI elements are made within a radius 4).

Like the MI sensory maps described above, the MI output map in residual intact cortex experienced an increase in relative activity. The number of MI elements activating one or more muscle group(s) increased following a MI lesion of sufficient size (4x4 and larger). For an 8x8 lesion, the percentage of remaining MI elements activating one or more muscle group(s) increased from 77% to 86% of intact elements. This affected the positioning of the model arm as well, when tested with six external inputs to MI. As seen in Figure B.1c, with a 16x16 focal lesion in MI the arm position revealed a weakened response immediately post-lesion. For example, the elbow position immediately post-lesion for the upper arm flexor test was 15° away from its pre-lesion position, roughly in the direction of the resting position. Further post-lesion synaptic modifications in the presence of the MI lesion did not produce a complete realignment of the arm positions with their pre-lesion location, although complete recovery did occur with smaller MI lesions (e.g., 8x8).

The lack of any significant post-lesion reorganization with small MI lesions (2x2 and 3x3) can be attributed to the coarseness of the topographic projections from PI to MI. Each MI element receives input from 61 PI elements, so with such small MI lesions the distribution of output from PI elements

was only minimally perturbed, and perilesion elements continued to experience a distribution of input patterns similar to that before lesioning. As a result their receptive fields, and thus the MI map, remained largely unchanged due to the correlational nature of the synaptic modification rule.

Examination of the feature maps for PI (both post-lesion and with further training) did not reveal any qualitative reorganization following MI lesions, beyond the small shifts of cluster positions expected with this model [Chen & Reggia, 1996]. While motor output was weakened with larger MI lesions, it did not appear to affect feature map organization in PI.

### B.3 Comments

This model demonstrates interesting post-lesion effects concerning cortical map reorganization, along with some insight into why these secondary effects arise. It was observed that focal lesions resulted in a two-phase map reorganization process in the intact perilesion cortical region. The first, very rapid phase was due to changes in activation dynamics, while the second, slow phase was due to synaptic plasticity. Thus, the model makes the prediction that biological perilesion map changes will be demonstrable within a few minutes of a cortical lesion. While there are a few experimental animal studies that have examined post-lesion cortical map reorganization (see below), none of these have measured maps immediately following the lesion. Recent experimental studies in animals have repeatedly shown map reorganization within minutes following focal deafferentation of cortex [Metzler & Marks, 1979; Gilbert & Wiesel, 1992]; our model predicts that they will occur following cortical lesions as well and provides some details about their nature.

The second prediction of our model is that increased perilesion excitability is necessary for effective map reorganization in cortex surrounding an acute focal lesion. When increased perilesion excitability was present during the first phase of map reorganization, the cortex surrounding the lesion consistently participated in the map reorganization process, even achieving a higher density feature map than in the prelesion cortex. Presumably such effective utilization of surrounding intact cortex following a lesion could contribute to behavioral recovery following an ischemic stroke. On the other hand, when there was decreased excitation in perilesion cortex, this intact cortex consistently did *not* participate in map reorganization, and the perilesion cortex that “dropped out” of the map actually expanded with time due to the normal modifications of synaptic strengths. These very different results, observed here for pure feature maps (PI) and for feature maps involving topographically arranged inputs (to MI from PI), are consistent with similar results obtained in our earlier study involving pure topographic maps [Sutton *et al.*, 1994; Armentrout *et al.*, 1994].

The notion that perilesion excitability is an important factor may prove useful in interpreting animal studies of post-lesion map reorganization. Under some conditions in these studies, functions originally represented in the infarct zone of sensorimotor cortex reappeared or expanded in nearby intact cortex [Jenkins & Merzenich, 1987; Nudo & Milliken, 1996; Castro-Alamancos & Borrel, 1995], while under other conditions they did not [Nudo *et al.*, 1996]. Our model suggests that assessing perilesion excitability under these differing conditions may shed light on why the different results occur.

The dependence of map reorganization upon perilesion excitability in the model can be explained by examining the synaptic modification rule that produces map formation originally. Informally, this rule causes changes to a cortical element’s receptive field 1) at a rate proportional to how active that element is, and 2) such that the receptive field shifts to become more like the pattern of input elements that activate that cortical element. Thus, when the activation of a perilesion element is

low, its receptive field changes very slowly and little reorganization occurs. When perilesion activity is high, the receptive field will change quickly and substantial reorganization will occur. In this context, the differences in the input connections to PI and MI account for differences in how these two regions reorganize. In PI, the diffuse afferent inputs have little influence on, and therefore little correlation with, the perilesion elements following a lesion. Thus intact cortical elements adjacent to the original post-lesion functional deficit lose correlated activity from neighbors, become less correlated with specific input patterns, and tend to drop out of the map. In contrast, the coarsely topographic connections from PI to MI that originally supply the outer region of lesioned cortex have an increased influence on, and become more correlated with, perilesion elements, causing the latter's receptive fields to shift and thus substantial map reorganization to occur.

In the context of these modeling results, it is interesting to note that there does exist direct experimental evidence for increased excitability in intact cortex following a small focal lesion [Domann *et al.*, 1995]. Such increased excitability has generally been viewed as detrimental, although this is controversial [Hossmann, 1994]. Our computational model suggests that, in addition, increased excitability may play an important and previously unrecognized role in recovery from stroke. At the very least, the model indicates that further experimental investigation of this issue is warranted and will be useful in obtaining a better understanding of recovery after stroke. In our model, the primary factors determining whether perilesion activity increased or decreased were the extent of divergence of afferents to the cortical region and the ratio of intracortical lateral excitation to inhibition. In other words, in both PI and MI the cortex immediately around the lesion lost excitatory input from the lesioned region. However, the widely divergent inputs to PI were insufficiently powerful to compensate for this loss of perilesion excitation from lateral connections arising in the lesion area, while the much more focused afferents to MI were.

## Bibliography

- [1] B. Angeniol, V. G. de La Croix, and J. Le Texier. Self-organizing feature maps and the traveling salesman problem. *Neural Networks*, 1:289–294, 1988.
- [2] S. L. Armentrout, J. A. Reggia, and Michael Weinrich. A neural model of cortical map reorganization following a focal lesion. *Artificial Intelligence in Medicine*, 6:383–400, 1994.
- [3] H. Asanuma. *The Motor Cortex*. Raven Press, 1989.
- [4] M. Benaim and M. Samuelides. Dynamical properties of neural nets using competitive activation mechanisms. In *Proceedings of International Joint Conference on Neural Networks*, volume 3, pages 541–546, 1990.
- [5] I. C. Bruce and W. G. Tatton. Sequential output-input maturation of kitten motor cortex. *Experimental Brain Research*, 39:411–419, 1980.
- [6] D. Bullock, S. Grossberg, and F. H. Guenther. A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm. *Journal of Cognitive Neuroscience*, 5(4):408–435, 1993.
- [7] Y. Burnod, P. Grandguillaume, I. Otto, S. Ferraina, P. B. Johnson, and R. Caminiti. Visuo-motor transformations underlying arm movements toward visual targets: A neural network model of cerebral cortical operations. *The Journal of Neuroscience*, 12(4):1435–1453, 1992.
- [8] M. Castro-Alamancos and J. Borrel. Functional recovery of forelimb response capacity after forelimb primary motor cortex damage. *Neuroscience*, 68:793–805, 1995.
- [9] B. Chapman, K. R. Zahs, and M. P. Stryker. Relation of cortical cell orientation selectivity to alignment of receptive fields of the geniculocortical afferents that arborize within a single orientation column. *The Journal of Neuroscience*, 11(5):1347–1365, 1991.
- [10] Y. Chen and J. Reggia. Alignment of coexisting cortical maps in a motor control model. *Neural Computation*, 8(4):731–755, 1996.
- [11] P. D. Cheney and E. E. Fetz. Comparable patterns of muscle facilitation evoked by individual corticomotoneuronal (cm) cells and by single intracortical microstimuli in primates: evidence for functional groups of cm cells. *The Journal of Neurophysiology*, 53:805–820, 1985.
- [12] S. Cho and J. A. Reggia. Map formation in proprioceptive cortex. *International Journal of Neural Systems*, 5:87–101, 1994.



- [13] S. Cho, J. A. Reggia, and C. L. D’Autrechy. Modeling map formation in proprioceptive cortex. Technical Report CS-TR-3026, University of Maryland, January 1993.
- [14] S. Cho, M. Jang, and J. Reggia. Effects of parameter variations on feature maps. In *Proceedings of International Conference on Neural Information Processing, Seoul, Korea*, pages 1301–1306, 1994.
- [15] Le Cun, Y. B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1:541–551, 1989.
- [16] R. Domann, G. Hagermann, M. Kraemer, H-J Freund, and O. Witte. Electrophysiological changes in the surrounding brain tissue of photochemically induced cortical infarcts in the rat. *Neuroscience Letters*, 155:69–72, 1995.
- [17] J. P. Donoghue, S. Leibovic, and J. N. Sanes. Organization of the forelimb area in squirrel monkey motor cortex: Representation of individual digit, wrist, and elbow muscles. *Experimental Brain Research*, 89(1):1–19, 1992.
- [18] E. Erwin, K. Obermayer, and K. Schulten. Self-organizing maps: ordering, convergence properties and energy functions. *Biological Cybernetics*, 67:47–55, 1992.
- [19] D. Felleman and D. Van Essen. Distributed hierarchical processing in primate cerebral cortex. *Cerebral Cortex*, 1:1–47, 1991.
- [20] W. H. Freeman, editor. *The Brain*, chapter 7, 8, 9. Scientific American, 1979.
- [21] S. Geffin and B. Furht. A dataflow multiprocessor system of robot arm control. *The International Journal of Robotics Research*, 9(3):93, 1990.
- [22] A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner. Neuronal population coding of movement direction. *Science*, 233:1416–1419, 1986.
- [23] C. Gilbert and T. Wiesel. Receptive field dynamics in adult primary visual cortex. *Nature*, 356:150–152, 1992.
- [24] C. Gilbert. Horizontal integration in the neocortex. *Trends in Neuroscience*, 8:160–165, 1985.
- [25] S. Goodall, J. Reggia, Y. Chen, E. Ruppin, and C. Whitney. A computational model of acute focal cortical lesions. *Stroke*, 28(1):101–109, 1997.
- [26] R. P. Gorman and T. J. Sejnowski. Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Networks*, 1:75–89, 1988.
- [27] K. Grajski and M. Merzenich. Hebb-type dynamics is sufficient to account for the inverse magnification rule in cortical somatotopy. *Neural Computation*, 2:71–84, 1990.
- [28] J. Hertz, A. Krogh, and R. G. Palmer. *Introduction to the Theory of Neural Computation*, chapter 5, 6, 9. Addison-Wesley, 1993.

- [29] R. Hess, K. Negishi, and O. Creutzfeldt. The horizontal spread of intracortical inhibition in the visual cortex. *Experimental Brain Research*, 22:415–419, 1975.
- [30] J. J. Hopfield and D. W. Tank. Neural computation of decisions in optimization problems. *Biological Cybernetics*, 233:625–633, 1985.
- [31] K. Hossmann. Glutamate-mediated injury in focal cerebral ischemia: the excitotoxin hypothesis revised. *Brain Pathology*, 4:23–26, 1994.
- [32] D. H. Hubel and T. N. Wiesel. Receptive fields of single neurones in the cat’s striate cortex. *Journal of Physiology*, 148:574–591, 1959.
- [33] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction, and functional architecture in the cat’s visual cortex. *Journal of Physiology*, 160:106–154, 1962.
- [34] D. H. Hubel and T. N. Wiesel. Receptive fields of cells in striate cortex of very young, visually unexperienced kittens. *Journal of Neurophysiology*, 26:994–1002, 1963.
- [35] W. Jenkins and M. Merzenich. Reorganization of neocortical representations after brain injury. In F. Sell, E. Herbert, and B. Carison, editors, *Progress in Brain Research*, pages 249–266. Elsevier, Amsterdam, Netherlands, 1987.
- [36] W. Jenkins, M. Merzenich, M. Ochs, T. Allard, and E. Guic-Robles. Functional reorganization of primary somatosensory cortex in adult owl monkeys after behaviorally controlled tactile stimulation. *Journal of Neurophysiology*, 63:229–231, 1990.
- [37] P. B. Johnson. Toward an understanding of the cerebral cortex and reaching movements: A review of recent approaches. In R. Caminiti, P. B. Johnson, and Y. Burnod, editors, *Control of Arm Movement in Space*. Springer-Verlag, 1992.
- [38] E. G. Jones, J. D. Coulter, and S. H. Hendry. Intracortical connectivity of architectonic fields in the somatic sensory, motor and parietal cortex of monkeys. *Journal of Computational Neurology*, 181:291–374, 1978.
- [39] E. R. Kandel and J. H. Schwartz, editors. *Principles of Neural Science*, chapter 33, 34, 38. Elsevier, 2nd edition, 1985.
- [40] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:41–65, 1982.
- [41] T. Kohonen. Representation of sensory information in self-organizing feature maps, and relation of these maps to distributed memroy networks. In Harold H. Szu, editor, *Optical and Hybrid Computing*. SPIE, 1987.
- [42] T. Kohonen. *Self-organization and Associative Memory*, chapter 3, 5. Springer-Verlag, 1989.
- [43] T. Kohonen. *Self-organizing Maps*. Springer-Verlag, 1995.
- [44] M. Kuperstein. Neural model of adaptive hand-eye coordination for single postures. *Science*, 239:1308–1311, 1988.

- [45] R. Linsker. From basic network principles to neural architectures. In *Proceedings of the National Academy of Sciences*, volume 83, pages 7508–7512, 8390–8394 and 8779–8783, 1986.
- [46] Z. Lo, Y. Yu, and B. Bavarian. Analysis of the convergence properties of topology preserving neural networks. *IEEE Trans. on Neural Networks*, 4(2):207–220, 1993.
- [47] T. Martinetz and K. Schulten. A “neural gas” network learns topologies. In *Proceedings of International Conference on Artificial Neural Networks, Vol.1*, pages 397–407, 1991.
- [48] B. W. Mel. Murphy: A robot that learns by doing. In *Neural Information Processing Systems*. American Institute of Physics, NY, 1988.
- [49] B. W. Mel, editor. *Connectionist Robot Motion Planning*. Academic Press, Inc., 1990.
- [50] J. Metzler and P. Marks. Functional changes in cat sensory-motor cortex during short-term reversible epidural blocks. *Brain Research*, 177:379–383, 1979.
- [51] K. D. Miller, J. B. Keller, and M. P. Stryker. Ocular dominance column development: Analysis and simulation. *Science*, 245:605–615, August 1989.
- [52] V Mountcastle. An organizing principle for cerebral function. In G. Edelman and V. Mountcastle, editors, *The Mindful Brain*. MIT Press, Cambridge, MA, 1978.
- [53] F. A. Mussa-Ivaldi, E. Bizzi, P. Morasso, and N. Hogan. Network models of motor systems with many degrees of freedom. In Martin D. Fraser, editor, *Advances in Control Network and Large-scale Parallel-distributed Processing Models*, volume 1, chapter 6. Ablex Pub. Corp, 1991.
- [54] S. Nicosia, P. Tomei, and A. Tornambi. Non-linear control and observation algorithm for a single-link flexible robot arm. *International Journal of Control*, 49(3):827, 1989.
- [55] R. Nudo and G. Milliken. Reorganization of movement representations in primary motor cortex following focal ischemic infarcts in adult squirrel monkeys. *Journal of Neurophysiology*, 75:2144–2149, 1996.
- [56] R. J. Nudo, B. M. Wide, F. Sifuentes, and G. W. Milliken. Neural substrates for the effects of rehabilitative training on motor recovery after ischemic infarct. *Science*, 272:1791–1794, 1996.
- [57] J. C. Pearson, L. H. Finkel, and G. M. Edelman. Plasticity in the organization of adult cerebral cortical maps: A computer simulation based on neuronal group selection. *The Journal of Neuroscience*, 7:4209–4223, December 1987.
- [58] W. Penfield and T Rasmussen. *The Cerebral Cortex of Man*. Macmillan, 1950.
- [59] L. L. Porter, T. Sakamoto, and H. Asanuma. Morphological and physiological identification of neurons in the cat motor cortex which receive direct input from the somatic sensory cortex. *Experimental Brain Research*, 80:209–212, 1990.
- [60] J. Ramanujam and P. Sadayappan. Optimization by neural networks. In *IEEE International Conference on Neural Networks*, volume 2, pages 325–332, 1988.

- [61] J. A. Reggia and M. Edwards. Phase transitions in connectionist models having rapidly varying connection strengths. *Neural Computation*, 2:523–535, 1990.
- [62] J. A. Reggia, C. L. D’Autrechy, G. G. Sutton, and M. Weinrich. A competitive distribution theory of neocortical dynamics. *Neural Computation*, 4:287–317, 1992.
- [63] J. Reggia, S. Goodall, Y. Chen, E. Ruppin, and C. Whitney. Modeling post-stroke cortical map reorganization. In J. Reggia, E. Ruppin, and R. Berndt, editors, *Neural Models of Brain and Cognitive Disorders*. World Scientific, 1996.
- [64] H. Ritter and K. Schulten. On the stationary state of kohonen’s self-organizing sensory mapping. *Biological Cybernetics*, 54:99–106, 1986.
- [65] H. Ritter, T. Martinetz, and K. Schulten. Topology-conserving maps for learning visuo-motor-coordination. *Neural Networks*, 2:159–168, 1989.
- [66] G. G. Sutton and J. A. Reggia. Effects of normalization constraints on competitive learning. *IEEE Transactions on Neural Networks*, 5(3):502–504, 1994.
- [67] G. G. Sutton, J. A. Reggia, C. L. D’Autrechy, and S. L. Armentrout. Cortical map reorganization as a competitive process. *Neural Computation*, 6:1–13, 1994.
- [68] S. Tanaka. Theory of ocular dominance column formation. *Biological Cybernetics*, 64:263–272, 1991.
- [69] T. J. Tarn, A. K. Bejczy, and X. Yun. Effect of motor dynamics on nonlinear feedback robot arm control. *IEEE Transactions on Robotics and Automation*, 7(1):114, 1991.
- [70] C. von der Malsburg. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14:85–100, June 1973.
- [71] J. A. Walter and Klaus J. Schulten. Implementation of self-organizing neural networks for visuo-motor control of an industrial robot. *IEEE Trans. Neural Networks*, 4:86–95, January 1993.
- [72] P. Y. Wang and S. B. Seidman. Analysis of competition-based spreading activation in connectionist models. *International Journal of Man-Machine Studies*, 28:77–97, 1988.
- [73] M. Weinrich, G. G. Sutton, J. A. Reggia, and C. L. D’Autrechy. Adaptation of noncompetitive and competitive neural networks to focal lesions. *Journal of Artificial Neural Networks*, 1:51–60, 1994.
- [74] T. G. Weyand, J. G. Malpeli, C. Lee, and H. D. Schwark. Cat area 17. iii. response properties and orientation anisotropies of corticotectal cells. *Journal of Neurophysiology*, 56:1088–1101, 1986.
- [75] D. A. White and D. A. Sofge, editors. *Handbook of Intelligent Control*, chapter 1. Multiscience Press, Inc., 1992.

- [76] S. Wise and J. Tanji. Neuronal responses in sensorimotor cortex to ramp displacement and maintained positions imposed on hindlimb of the unanesthetized monkey. *Journal of Neurophysiology*, 45:482–500, 1981.
- [77] H. Yumiya and C. Ghez. Specialized subregions in the cat motor cortex. *Experimental Brain Research*, 53:259–276, 1984.