CHAPTER 6

SPATIAL MAPS OF MOTOR PATTERNS

6.1. The General Problem: Transforming Pattern Intensities into Map Positions

The problem of this chapter is to analyse how different activity patterns over a fixed set of cells can be transformed into different spatial foci of activity at the next processing stage. We call a network which encodes information by activating cells in different locations a spatial map. Our problem is thus to transform different activity patterns over a fixed set of cells into a spatial map. In the case of the HMI, the activity patterns represent difference vectors. Hence the spatial map is consistent with retinotopic coordinates (Section 4.8). We therefore call such a spatial map a retinotopic map, or RM. As we noted in Section 5.11, the spatial map that recodes tonic cell outflow patterns is called an eye position map, or EPM. Such an EPM is consistent with head coordinates, rather than retinotopic coordinates.

Although an RM and an EPM encode different types of information, they can arise from the same mechanisms. The formal problem to be solved is the same, no matter what interpretation is given to the activity patterns that must be spatially parsed.

Several types of mechanisms are consistent with these functional requirements. The available neural data do not unambiguously force the choice of one mechanism above all others. Evolutionary variations could, moreover, choose different solutions across species. We therefore analyse several possible spatial mapping models, much as we described several possible models of saccadic error correction in Section 3.18. Each of these models leads to testable differences that future experiments can attempt to measure, and the collection of all the models provides a deeper conceptual insight than any one model could provide.

To fix ideas, we will henceforth discuss the special cases of mapping HMI vectors into an RM, or of T cell outflow patterns into an EPM. The same considerations hold for transforming any set of muscle-coded activity patterns into a spatial map.

6.2. Antagonistic Positional Gradients, Contrast Enhancement, and Coincidence Detectors

The first model executes a transformation that requires no learning. It exploits the organization of muscles into agonist-antagonist pairs. A similar construction exists for sensory fields that possess natural left-to-right and/or bottom-to-top symmetry axes. A similar map seems, for example, to exist in the auditory system of the owl (Konishi, 1984).

The model's inputs are grouped into agonist and antagonist pairs. Both inputs need to be positive, except in the extremal cases where one
input is maximally active and the other is shut off. In the MLN, both agonist and antagonist tonic cells can directly give rise to positive inputs. By contrast, only one output from each agonist-antagonist pair of the HMI can be positive. Hence such outputs need to activate tonically active cells in a push-pull fashion, as in equation (5.15), before these tonic cells can generate inputs to the model.

Denote the input pair by \((I_1, I_2)\). The model first maps each input pattern into a one-dimensional spatial map in such a way that different map positions correspond to different input ratios \(I_1/I_2\). To see how this happens, denote the population with input activity \(I_i\) by \(v_i\), \(i = 1, 2\). Let each population \(v_i\) send pathways to a field \(F_i\) of cells. Since there are many cells in \(F_1\), we approximate the field of cells by a continuous one-dimensional medium. Let \(S\) be the spatial variable of this medium, and let \(P_{iS}\) denote the strength of the pathway from \(v_i\) to position \(S\) in \(F_i\). Suppose for definiteness that

\[
P_{1S} = -P e^{-\mu(v-S)^2} \tag{6.1}
\]

and

\[
P_{2S} = P e^{-\mu(\omega-S)^2} \tag{6.2}
\]

In other words, each cell population \(v_i\) sends a broad spatial gradient of connections to \(F_i\) (Figure 6.1). Each spatial gradient connects to \(F_i\) according to a simple random growth law. By equation (6.1), the best connection of \(v_1\) is to position \(S = \nu\). Population \(v_1\) contacts other positions \(S\) within \(F_1\) with a strength that decreases as a Gaussian function of their distance \(|\nu - S|\) from position \(\nu\). By equation (6.2), the best connection of \(v_2\) is to position \(S = \omega\). Population \(v_2\) contacts other positions \(S\) within \(F_1\) with a strength that decreases as a Gaussian function of their distance \(|\omega - S|\) from position \(\omega\). Suppose that \(\nu < \omega\). The connections from \(v_1\) to \(F_1\) are assumed to be inhibitory. The connections from \(v_2\) to \(F_2\) are assumed to be excitatory. The total input at position \(S\) due to the activity pattern \((I_1, I_2)\) is

\[
J(S; I_1, I_2) = -I_1 P_{1S} + I_2 P_{2S}. \tag{6.3}
\]

We denote by \(S(I_1/I_2)\) the position \(S\) which receives the maximal input in response to input pattern \((I_1, I_2)\). As the input pattern \((I_1, I_2)\) changes, the position \(S(I_1/I_2)\) changes too, and depends only on the ratio \(I_1/I_2\). Thus the antagonistic interaction of a pair of Gaussian positional gradients can convert spatial patterns into spatial maps. Before proving this fact, we summarize the subsequent two stages of the model.

The input pattern to the field \(F_1\) is contrast-enhanced and normalized by on-center off-surround interactions within \(F_1\) (Section 2.6). Thus each activity pattern \((I_1, I_2)\) activates a sharply tuned population of cells within \(F_1\) at and near position \(S(I_1/I_2)\). This contrast-enhancement operation converts \(F_1\) into a one-dimensional spatial map of the agonist-antagonist
Figure 6.1. Mapping an agonist-antagonist pair \((v_1, v_2)\) of cell populations into a spatial map \(F_1\): In response to the input pattern \((I_1, I_2)\), the maximally activated position is \(S(I_1/I_2)\), which depends only upon the ratio of \(I_1\) to \(I_2\). Population \(v_1\) gives rise to a broadly distributed inhibitory spatial gradient. Population \(v_2\) gives rise to a broadly distributed excitatory gradient.
input patterns \((I_1, I_2)\). [A similar organization helps to explain the peak shift and behavioral contrast that can occur during discrimination learning experiments (Grossberg, 1981).

We now consider how to generate a spatial map of more than one agonist-antagonist activity pattern. In cases where just two agonist-antagonist patterns are needed, this can be done by letting each location in \(F_1\) activate parallel strips of cells at the next stage \(F_2\) (Figure 6.2). Assume that the other agonist-antagonist pair also activates parallel strips of cells in \(F_2\), but that the strips of the different agonist-antagonist pairs are not parallel. The ideal case is one in which the two types of strips are mutually perpendicular. Finally, suppose that the cells within \(F_2\) are coincidence detectors that respond only if they are simultaneously activated by a pair of strips. Such coincidence detectors are easily designed due to the fact that the inputs from \(F_1\) to \(F_2\) are normalized.

Such a field \(F_2\) computes a two-dimensional spatial map of the two pairs of agonist-antagonist input patterns. This procedure can be iterated to generate an \(n\)-dimensional spatial map, but it becomes physiologically implausible for values of \(n\) much larger than 3.

It remains to determine the position \(S(I_1/I_2)\) at which the input pattern \(J(S; I_1, I_2)\) of equation (6.3) is maximal. To locate this position, we determine the solutions of the equation

\[
\frac{\partial}{\partial S} J(S; I_1, I_2) = 0. \tag{6.4}
\]

Equation (6.4) implies that

\[
\frac{(S - \omega)}{(S - \nu)} e^{((S - \nu)^2 - (S - \omega)^2)} \frac{I_1}{I_2}. \tag{6.5}
\]

Both \(I_1/I_2\) and the exponential term are nonnegative. Hence (6.5) has a solution only if the ratio \((S - \omega)(S - \nu)^{-1}\) is also nonnegative. This is true only if \(S \leq \nu\) or \(S \geq \omega\). Since \(\nu < \omega\) and the connections from \(v_1\) to the cells around position \(S = \nu\) are inhibitory, it is clear that some nonnegative inputs \(J(S; I_1, I_2)\) are found in the region \(S \geq \omega\), which we consider henceforth.

Introducing the new variable \(y = S - \nu\), we rewrite (6.5) in the form

\[
\frac{(y - \lambda)}{y} e^{y [(\lambda - y)^2]} \frac{I_1}{I_2},
\]

where \(\lambda = \omega - \nu\). Equation (6.6), in turn, implies that

\[
f(y) = \frac{e^{\mu \lambda^2} I_1}{I_2}
\]
Figure 6.2. Coincidence detectors at level $F_2$ selectively respond to pairs of spatial positions at $F_1$ which are activated by different agonist-antagonist cell populations.
where

\[ f(y) = \frac{(y - \lambda)}{y}e^{2\mu\lambda y}. \]  

(6.8)

Since \( y - \nu = S \geq \omega \), it follows that \( y \geq \lambda \). As \( y \) increases from \( \lambda \) to \( \infty \), \( f(y) \) increases from 0 to \( \infty \). Thus given any nonnegative choice of the ratio \( I_1/I_2 \), there exists a value \( y(I_1/I_2) \) of \( y \) between \( \lambda \) and \( \infty \) that satisfies (6.7). Moreover, as the ratio \( I_1/I_2 \) increases, the value of \( y(I_1/I_2) \) also increases. In terms of the positional variable \( S = y + \nu \), this means that the position \( S(I_1/I_2) \) at which \( J(S; I_1, I_2) \) is maximal increases from \( \omega \) to \( \infty \) as the ratio \( I_1/I_2 \) increases from 0 to \( \infty \).

Of course, it makes no sense to talk about an infinite field of cells. Thus it is necessary to consider how the maximal range of input ratios \( I_1/I_2 \) can be coded by a finite interval of cells. Inspection of (6.7) and (6.8) shows that a choice of parameters such that \( \lambda \) is small and \( \mu \lambda \) is sufficiently large facilitates this goal. Then the function \( (y - \lambda)y^{-1} \) in (6.8) rapidly jumps from 0 towards its maximum value 1 as \( y \) increases above \( \lambda \). Function \( e^{2\mu\lambda y} \) in (6.8) grows quickly as a function of \( y \), thereby being able to match a larger range of ratios \( I_1/I_2 \) in (6.7). Finally the coefficient \( \mu \lambda^2 \) in (6.7) can be relatively small even though \( \mu \lambda \) is relatively large, because \( \lambda \) is small. Thus the main constraint is that the amount of shift \( \lambda = \omega - \nu \) of the antagonistic positional gradients should be small relative to other parameters of the problem. Increasing the spatial decay rate \( \mu \) of these spatial gradients enables a wider range of input ratios \( I_1/I_2 \) to be encoded within a fixed interval.

### 6.3. Position-Threshold-Slope Shift Maps

The previous model illustrated the importance of using antagonistic quantities to convert activities into positions. The next model describes a different version of this general idea. In this model, the antagonism is not derived from the interaction of an agonist-antagonist pair. The antagonism arises from a single motor activity. Using such a model, the HMI does not have to preprocess its outputs by putting them through a tonic push-pull process. To fix ideas, we consider how such a model would convert the vector differences \( V = ([x_1]^+, [x_2]^+, \ldots, [x_6]^+) \) of the HMI into activated positions within the RM.

Using this model, each suprathreshold activity \( [x_j]^+ \) at the HMI generates a signal \( [x_j]^+P_j(r, \theta) \) to a position \( (r, \theta) \) at the first stage \( F_1 \) of processing, which is depicted for convenience as a surface in polar coordinates \( (r, \theta) \). The path strengths \( P_j(r, \theta) \) define positional gradients from each HMI population \( v_j \) to \( F_1 \). The total input from the HMI to position \( (r, \theta) \) of \( F_1 \) is the sum

\[ S(r, \theta) = \sum_{j=1}^{6}[x_j]^+P_j(r, \theta) \]
Spatial Maps of Motor Patterns

of these signals (Figure 6.3). Then an output signal

$$T(r, \theta) = \left[ S(r, \theta) - \Gamma(r, \theta) \right]^+$$  \hspace{1cm} (6.10)

is emitted to the second stage $F_2$ of processing, which is the RM. Recall that the notation $[\xi]^+$ stands for $\max(\xi, 0)$.

By (6.9), the input $S(r, \theta)$ to $F_1$ is the sum of output signals $[x_j]^+$ weighted by the path strengths $P_j(r, \theta)$. In other words, the HMI vector $V = ([x_1]^+, [x_2]^+, \ldots, [x_6]^+)$ is filtered by the path strength vector $P(r, \theta) = (P_1(r, \theta), P_2(r, \theta), \ldots, P_6(r, \theta))$ at position $(r, \theta)$. For simplicity, we have assumed that $F_1$ equilibrates rapidly to this input, and that the equilibrium activity at each position $(r, \theta)$ equals its input $S(r, \theta)$. Then $S(r, \theta)$ gives rise to the output signal $T(r, \theta)$ in (6.10). The map $S(r, \theta) \rightarrow T(r, \theta)$ from $F_1$ to the RM is topographic, whereas the map $V \rightarrow S(r, \theta)$ from the HMI to $F_1$ is built up from the convergent filtering action of six positional gradients.

Equation (6.10) says that a signal is emitted to the RM only if $S(r, \theta) > \Gamma(r, \theta)$. Thus $\Gamma(r, \theta)$ is the signal threshold of the $(r, \theta)$ pathway. The signal $T(r, \theta)$ grows linearly as a function of suprathreshold values of $S(r, \theta)$, as in the classical Hartline-Ratliff equation (Ratliff, 1965) and other models of neural pattern discrimination (Grossberg, 1970, 1976a).

Both the positional gradients $P_j(r, \theta)$ and the signal thresholds $\Gamma(r, \theta)$ depend upon the position $(r, \theta)$ within $F_1$. A larger choice of $P_j(r, \theta)$ says that an input $[x_j]^+P_j(r, \theta)$ to $(r, \theta)$ grows with a steeper slope as a function of $[x_j]^+$. A larger choice of $\Gamma(r, \theta)$ says that a larger total input $S(r, \theta)$ to $(r, \theta)$ is needed to fire a signal to the RM. Both $P_j(r, \theta)$ and $\Gamma(r, \theta)$ are assumed to increase with $r$. Thus cells at $F_1$ with larger radial positions $r$ are more sensitive to their inputs and have higher output thresholds than cells with smaller radial positions $r$. This covariation of cell position, signal threshold, and signal slope causes a shift to occur in the spatial locus of maximal total activity at the RM as an HMI signal $[x_j]^+$ increases. Low intensity inputs cause maximal activity to occur at the low threshold end of the RM, whereas high intensity inputs cause maximal activity to occur at the high threshold end of the RM. Due to the role of correlations in cell position, threshold, and slope in generating this shift in activity locus, we call this mechanism a **Position-Threshold-Slope (PTS) Shift**. Populations of cells in which threshold and slope covary across cells have been found, for example, in the abducens and oculomotor nuclei (Luschei and Fuchs, 1972; Robinson, 1970; Schiller, 1970).

The input pattern to the RM that is caused by a PTS shift is contrast-enhanced and normalized before it is stored in STM as an RM activity peak. This is accomplished by endowing the RM with a suitably designed recurrent on-center off-surround shunting network (Section 2.6). A simple...
Figure 6.3. A position-threshold-slope (PTS) shift: Spatial patterns of activity within an HMI, or other source of spatial patterns, give rise to broadly distributed outputs that converge upon the network level $F_1$. Level $F_1$ maps topographically to level $F_2$. Due to the parameters of the broad spatial interactions and the topographic mapping, different spatial patterns activate different positions at $F_2$. The inputs $S$ and outputs $T$ are defined by equations (6.9) and (6.10).
Spatial Maps of Motor Patterns

rule that approximates the dynamics of such a network is

\[ x(r, \theta) = \left( \frac{T(r, \theta)}{\max(R, \phi) T(R, \phi)} \right)^n, \tag{6.11} \]

where \( n > 1 \). By (6.11), all nonmaximal \( T(r, \theta) \) generate a small \( x(r, \theta) \) if \( n \) is chosen sufficiently large, whereas the maximal \( T(r, \theta) \) generates an \( x(r, \theta) = 1 \) no matter how large \( n \) is chosen.

A 3-dimensional (one dimension for every agonist-antagonist pair) PTS shift can be designed so that every realizable HMI pattern \( V \) generates a different activity peak within the RM. Moreover, continuous variations in one or more activities within \( V \) cause continuous changes in the RM position. This PTS shift mechanism works well formally and uses known types of parametric correlations within neuronal populations. The most demanding constraint concerns the spatial juxtaposition of the six populations that respond to the six signals \( \{X_j\}^+ \) in \( V, j = 1, 2, \ldots, 6 \). Each population corresponds to a different muscle of one eye. In order for the PTS shift mechanism to work well, all six populations must be topographically placed in such a way that any three contiguous eye muscles are represented by three contiguous populations. The simplest way to accomplish this is to arrange the populations as a sector map (Figure 2.2).

Figures 6.4 and 6.5 summarize a computer simulation of how the RM changes position as a function of \( V \). We now describe the parameters used in this simulation. This description is rather technical, and can be skipped if the reader wishes to go to the next issue.

In any realizable monocular pattern \( V \) within the HMI, at most three consecutive components can be positive at any time. This is true because, if three contiguous muscles contract to saccade the eye, then their three antagonist muscles must relax. By “consecutive” components, we mean “consecutive modulo 6,” so that positive components \((x_5, x_6, x_1)\) and \((x_6, x_1, x_2)\) are considered consecutive. If we consider the set of all realizable patterns \( V \), then we can group them in terms of patterns which correspond to saccades in the same direction. When we do so, a convenient representation of the positive pattern components becomes apparent.

For example, consider the direction and length of the saccade encoded by a prescribed pattern \( V \). Suppose that pattern \( V^* \) encodes a saccade in the same direction, but of greater length. Then the same muscles that contract in response to \( V \) also contract in response to \( V^* \), but each contracting muscle must contract more in response to \( V^* \). In other words, each of the positive components in \( V^* \) is larger than its corresponding component in \( V \). This argument constrains the possible combinations of positive entries that can realize patterns \( V \).

We can represent patterns \( V \) whose saccades have increasing length but the same direction as triples of points on a succession of nonintersecting, expanding, closed curves. The distance of a point from the origin represents the size of a positive \( x_i \) in \( V \). The direction of the point with
Figure: Computer simulation of transform from HMI into map RM using TS shift. See text for details.
Figure 6.5. Computer simulation of a transformation from an HMI in (a) to an RM in (b). The parameters $(A, D)$ and $(r, \theta)$ which define the coordinates of the transformation are defined in equations (6.12)-(6.14). Due to the radial symmetry of the map, a similar transformation holds in all four quadrants. The similarity of the polar grid in $(r, \theta)$ coordinates to that in $(A, D)$ coordinates says that the map is approximately linear.
respect to the horizontal axis represents the direction in which its corresponding muscle contracts. This observation can be expressed more formally as follows. We represent each HMI output in the form

\[ [z_i]^+ = A f(|D - D_i|), \quad (6.12) \]

\( i = 1, 2, \ldots, 6. \) In (6.12), the parameters \( D_i \) and the function \( f(w) \) are fixed. They embody the constraints on saccadic length and direction that we have just summarized. The variables \( A \) and \( D \) change as a function of \( V \). Variable \( A \) encodes amplitude information and variable \( D \) encodes directional information. Thus the representation (6.12) expresses the six-dimensional vector \( V \) as a two-dimensional vector \((A, D)\) in polar coordinates. This is possible due to the manner in which the parameters \( D_i \) and the function \( f(w) \) absorb the extra degrees of freedom.

Parameter \( D_i \) represents the \( i \)th direction of contraction, \( i = 1, 2, \ldots, 6. \) Each pattern \( V \) defines a value \( A \) that increases with the length of the saccade. Function \( f(w) \) modulates the amplitude of \( A \) in a manner that depends upon which direction is under consideration. In particular, function \( f(w) \) has the following properties:

i) \( f(w) > 0 \) if \(-\pi/2 < w < \pi/2\);
ii) \( f(w) = 0 \) if \(-\pi \leq w \leq -\pi/2 \) or if \( \pi/2 \leq w \leq \pi \);
iii) \( f(w) \) increases if \(-\pi/2 < w < 0\);
iv) \( f(w) \) decreases if \( 0 < w < \pi/2\);
v) \( f(w) \) is \( 2\pi \)-periodic.

Using function \( f(w) \), each pattern \( V \) defines a value \( D \) that represents the direction of the saccade. For example, if \( D = D_3 \), then the saccade moves in a direction close to that represented by the third muscle. This follows from properties (iii) and (iv), since then \( x_3 = A f(0) \), whereas \( x_2 = A f(|D_3 - D_2|) < A f(0) \) and \( x_4 = A f(|D_3 - D_4|) < A f(0) \). Moreover at most three contiguous \( x_i \)’s are positive. This is because it is assumed that \(|D_3 - D_5| > \pi/2\), \(|D_5 - D_6| > \pi/2\), \(|D_3 - D_1| > \pi/2\), and so on. Thus by property (ii), if \( D = D_3 \) then \( x_1 = x_5 = x_6 = 0 \). A similar analysis holds for any choice of \( D \) such that \(-\pi \leq D \leq \pi\).

The PTS shift hypothesis implies that both the path strength \( P_i(r, \theta) \) and the threshold \( \Gamma(r, \theta) \) are increasing functions of \( r \). The gradient nature of the HMI→RM map suggests, in addition, that \( P_i(r, \theta) \) decreases as \( \theta \) deviates from the direction represented by \( D_i \). In our simulations, the thresholds \( \Gamma(r, \theta) \) were chosen independent of \( \theta \) for simplicity, but an increase of \( \Gamma(r, \theta) \) as \( \theta \) deviates from \( D_i \) is also physically plausible.

We numerically analysed several choices of the functions \( f(w) \), \( P_i(r, \theta) \), and \( \Gamma(r, \theta) \) and the parameters \( D_i \) in order to understand the PTS shift map. Our results indicated that a certain amount of regularity in these functions as \( i, r, \) and \( \theta \) vary generates a more uniformly distributed mapping function; that is, a mapping function whose chosen position within
the RM does not change very slowly as a function of certain changes in $V$
and very quickly as a function of other changes in $V$. Map uniformity is
not, however, an end in itself, since all that is needed is an (approximately)
one-to-one map.

The essentially linear map described in Figures 6.4 and 6.5 was generated using the following functions. First the $[x_i]^+$ in (6.12) were expressed in terms of the functions $D_i = \frac{(i-1)\pi}{3}$, $i = 1, 2, \ldots, 6$, and $f(w) = [\cos w]^+$. Then the PTS shift was defined using the positional gradients

$$P_i(r, \theta) = r[\cos(\theta - D_i)]^+, \tag{6.13}$$

in (6.9), and the signal thresholds

$$\Gamma(r, \theta) = \gamma r^2 \tag{6.14}$$
in (6.10), where $\gamma = .8$.

6.4. Self-Organizing Spatial Maps

The design of a 3-dimensional PTS shift map requires careful preprocessing of the HMI output vector, say via a sector map. An alternative solution to the spatial mapping problem does not need to impose this topographic constraint. In this solution, each input $[x_j]^+$ is again used to generate a PTS shift, but only within its own population. Distinct populations do not have to be embedded within a sector map. The output signals from these independent PTS shifts are the inputs to an adaptive coding model. This model for developmental map formation has also been used to explain several other types of data (Grossberg, 1976a, 1976b, 1982a). Thus, in this solution, a carefully, but plausibly, prewired network topography is replaced by a less carefully prewired network that can develop its own topography. Consequently, this self-organizing model can handle input vectors of any dimension.

The adaptive filter executes the same types of computations that were used to calibrate the TPM→HMI transform. Once again, the component signals $[x_j]^+$ in a vector input pattern such as $V$ are the starting point of the computation. The signals $[x_j]^+$ cannot themselves be used as the inputs to the adaptive filter from the HMI to the RM because $V$ encodes information about saccade direction and length. Why this is so is explained in later paragraphs. Instead, the intensities $[x_j]^+$ are first converted into spatial maps using one-dimensional PTS shifts. Then the total activity of all the active PTS shifts is normalized, or conserved, by long-range shunting lateral inhibition. Finally, the normalized positional activities $S_j$ corresponding to each $[x_j]^+$ are used as the inputs to the adaptive filter. The adaptive filter interacts with the RM to establish a spatial map of patterns $V$ within the HMI (Figure 6.6). We now describe this self-organizing spatial map in greater detail.
Figure 6.6. A self-organizing spatial map: Activity patterns within an HMI, or other source of spatial patterns, map into 1-dimensional PTS shifts, whose activities are normalized within $F_1$. Then $F_1$ generates output signals to $F_2$ via an adaptive filter. See text for details.
In a one-dimensional PTS shift, each intensity \([x_j]^+\) gives rise to an input that is uniformly distributed (in a statistical sense) across a population of cells. The signal thresholds and suprathreshold sensitivities (slopes) of the cells within each population are nonuniformly distributed within the population. Cells with lower thresholds tend to be found at one end of the population, cells with higher thresholds are found at the other end of the population, and cells with intermediate thresholds are found in between the two ends. Cells with higher thresholds are more sensitive to suprathreshold inputs. In other words, if the threshold of the cell's signal function is higher, then the slope of the signal function is steeper in the suprathreshold range. Thus the equation

\[
T_j(r) = \left[[x_j]^+ - \Gamma_j(r)\right]^+ P_j(r) \tag{6.15}
\]

defines the output from the \(j\)th muscle coordinate of the HMI to the radial position \(r\) of the \(j\)th PTS shift population. Since \(\Gamma_j(r) \geq 0\), (6.15) can be simplified to read

\[
T_j(r) = \left[x_j - \Gamma_j(r)\right]^+ P_j(r), \tag{6.16}
\]

which shows that the \(j\)th output can be computed directly from the HMI potential \(x_j\). In (6.16), both the threshold \(\Gamma_j(r)\) and the path strength \(P_j(r)\) are assumed to increase with \(r\). The largest signal \(T_j(r)\), across all values of \(r\), defines the PTS position activated by \(x_j\).

Denote by \(F_1\) the stage that receives the output signals (6.16) from all the HMI potentials \(x_j\). Thus \(F_1\) contains all the separate PTS shifts. We assume that \(F_1\) normalizes its response to this total input and stored it in STM. Denote the normalized STM activities of \(F_1\) by \(S_i\). Keep in mind that \(i \neq j\) because a whole population of cells corresponds to each HMI potential \(x_j\). These normalized activities \(S_i\) are the inputs to the adaptive coding model.

Each output signal \(S_i\) from \(F_1\) generates an input signal \(S_i C_{ik} z_{ik}\) to each population \(v_k\) of the spatial map. Parameter \(C_{ik}\) is the strength of the path that carries \(S_i\) to \(v_k\), and \(z_{ik}\) is the LTM trace of this path. For simplicity, suppose that all \(C_{ik} = 1\). This choice implies that the spatial gradients from the HMI to the RM are nonspecifically distributed to all RM populations. Then the signal due to \(S_i\) at \(v_k\) is \(S_i z_{ik}\). All of these signals are added to generate the total input

\[
T_k = \sum S_i z_{ik} \tag{6.17}
\]

that is received by \(v_k\).

Suppose that the RM responds to these inputs by satisfying the following properties:

1. It enhances the potentials \(x_k\) in the RM that correspond to the largest inputs \(T_k\) and suppresses the activities of all other potentials.
2. It conserves, or normalizes, the total activity of the enhanced potentials.

3. It stores the normalized and enhanced activities $x_k$ in STM until the RM is attentionally reset by auxiliary signals.

Letting $x_k$ be the activity of $V_k$, these rules can be approximated by the computation

$$x_k = \begin{cases} 1 & \text{if } T_k > \max\{\epsilon, T_m : m \neq k\} \\ 0 & \text{otherwise,} \end{cases} \quad (6.18)$$

which is a variant of (6.11).

6.5. Activity-Dependent Map Formation

Then the enhanced activities $x_k$, and only these activities, change the weights of the LTM traces $z_{ik}$. This adaptive coding postulate assumes that code learning is dependent upon post-synaptic activation. Singer (1983) has reported compatible data in the visual cortex. Due to (6.18), this postulate implies that learning occurs at the LTM trace $z_{ik}$ only if $x_k = 1$; that is, only if $V_k$ is activated. The role of the learning rule is to guarantee that the pattern $z_k \equiv (z_{1k}, z_{2k}, z_{3k}, \ldots)$ of LTM traces abutting $V_k$ becomes parallel, or proportional, to a time average of all the signal patterns $S \equiv (S_1, S_2, S_3, \ldots)$ that are active at $F_1$ when $V_k$ is active in the RM. If $V_k$ is active only when a single signal pattern $S$ is active, then $z_k$ becomes proportional to only this signal pattern. The simplest possibility is thus that $z_{ik}$ approaches $S_i$ as learning proceeds. If a larger set of signal patterns is active when $V_k$ is active, then the tuning curve of $V_k$ will become coarser, other things being equal.

The simplest learning law that embodies these properties is

$$\frac{d}{dt} z_{ik} = (-B z_{ik} + S_i)[x_k]^+. \quad (6.19)$$

Equation (6.19) is conceptually the same learning law as equation (4.2) within the HMI. Term $S_i$ in (6.19) plays the role of $S_i$ in (4.2), and term $[x_k]^+$ in (6.19) plays the role of both $P$ and $[x_j]^+$ in (4.2).

When a series of vector patterns $V$ sends signals from the HMI to the RM through time, the interaction of fast contrast enhancement and normalization in the RM with slow LTM adaptation parses the patterns $V$ in such a way that different patterns across the HMI activate topographically distinct regions of the RM. Many refinements of this mechanism can be implemented. For example, the prewired gradients $C_{ik}$ in (6.18) can be chosen so that strong agonist activations in the HMI tend to activate a given region of the RM, whereas strong antagonist activations in the HMI tend to activate a complementary region of the RM. For present purposes, the most important property of this model is that it is temporally stable. That is, because a small number of populations in $F_1$ is mapped into a
large number of populations in the RM, once a topographic parsing by the RM of patterns across the HMI is established, this spatial map will endure through time unless the system suffers some type of internal damage.

We will now indicate in greater detail how the rules (6.17)-(6.19) work together to control spatial map formation within an RM or EPM, and along the way explain why the output from the HMI or tonic cells must be fed into PTS shifts before the normalized PTS shift response generates inputs to the adaptive coding model.

Since, by (6.18), \( x_k = 0 \) or \( 1 \), (6.19) can be rewritten as

\[
\frac{d}{dt} z_{ik} = (-Bz_{ik} + S_i)x_k.
\]

(6.20)

For notational simplicity, also let \( B = 1 \). By (6.20), \( x_k = 0 \) implies that \( \frac{d}{dt} z_{ik} = 0 \). Consequently, no learning occurs in \( z_{ik} \) unless \( v_k \) is activated.

If \( x_k = 1 \), then (6.20) implies

\[
\frac{d}{dt} z_{ik} = -z_{ik} + S_i.
\]

(6.21)

By (6.21), \( z_{ik} \) approaches \( S_i \) whenever \( v_k \) is active. In other words, the LTM pattern \( z_k \) approaches the signal pattern \( S \) as learning proceeds. This property holds for any \( S \) that is active when \( v_k \) is active. Thus, integrating equation (6.21) shows that \( z_k(t) \) is a time average of all the signal patterns \( S \) that were active when \( v_k \) was active before time \( t \).

The definition in (6.17) of the total input \( T_k \) to \( v_k \) shows why training each \( z_k \) to become parallel to a different signal pattern \( S^{(k)} = (S_1^{(k)}, S_2^{(k)}, S_3^{(k)}, \ldots) \) enables each \( S^{(k)} \) to selectively activate its \( v_k \). Under these circumstances, in response to any signal pattern \( S \), \( T_k = \sum_i S_i S_i^{(k)} \) and \( T_m = \sum_i S_i S_i^{(m)} \) for all \( m \neq k \). If we choose \( S = S^{(k)} \), then \( T_k = \sum_i [S_i^{(k)}]^2 \) whereas \( T_m = \sum_i S_i^{(k)} S_i^{(m)} \) for all \( m \neq k \). Other things being equal, \( T_k > T_m \), \( m \neq k \), because \( S^{(k)} \) is parallel to itself in \( T_k \) but \( S^{(k)} \) is not parallel to any of the other signal patterns \( S^{(m)} \) in \( T_m \), \( m \neq k \). Whenever \( T_k > T_m \), \( m \neq k \), it follows by (6.18) that \( x_k = 1 \) and all \( x_m = 0 \), \( m \neq k \).

In other words, \( S^{(k)} \) is coded by \( v_k \).

In order to guarantee that \( T_k > T_m \) whenever \( z_k = S^{(k)} \) and \( z_k \neq S^{(m)} \), \( m \neq k \), one problem still needs to be overcome. The total signal \( \sum_i S_i \) needs to be constant (conserved, normalized) across all signal patterns \( S \). Otherwise, the encoding mechanism can fail as follows. Let \( z_k = S^{(k)} \) and \( z_m = S^{(m)} \), \( m \neq k \), but suppose that each signal \( S_i^{(m)} \) in \( S^{(m)} \) is larger than the corresponding signal \( S_i^{(k)} \) in \( S^{(k)} \). Then \( \sum_i S_i^{(m)} > \sum_i S_i^{(k)} \), so the total signal is not normalized. In this situation, \( T_m \) can be larger than
Tk even if $S = S^{(k)}$ because, although $z_k$ is parallel to $S$, each LTM trace in $z_m$ is larger than the corresponding LTM trace in $z_k$. The total activity across $F_1$ was normalized to deal with this problem.

6.6. Coding of Movement Length and Direction

In many coding problems, normalizing the input pattern itself causes no difficulty. If, however, we directly normalized the signals $[x_j]^+$ from the HMI and fed these normalized signals into the adaptive filter, then a serious problem would have been caused. The patterns $V$ at the HMI represent saccade direction and length. Normalizing these patterns would collapse all saccadic commands with the same direction into a single normalized command $S$. Length differences of these saccades could not be distinguished by this mechanism.

This problem is solved by letting each $[x_j]^+$ input to a 1-dimensional PTS shift. This transformation maps HMI intensities into PTS positions in an approximately one-to-one fashion. Normalizing the positive intensities of the activated PTS positions does not alter the one-to-one property of the map. No loss of saccadic length or direction information occurs if normalization acts on output signals from PTS shifts. Then the normalized PTS shift pattern $S$ can be fed as input signals to the adaptive coding model to generate a spatial parsing of HMI patterns across the RM.

6.7. Normalization of Total PTS Shift Map

It remains to say how the total output from all the PTS shifts can be normalized. The physiologically simplest schemes feed the PTS intensities as inputs into an on-center off-surround network whose cells obey membrane equations (shunting interactions) and whose off-surrounds are broadly distributed across the network. A network with particularly nice properties undergoes both feedforward and feedback on-center off-surround interactions. The inputs and the feedback signals activate the same populations of interneurons, which distribute both types of signals to the rest of the network. The feedback signals are chosen to be linear functions of population activity. The choice of linear feedback signals may cause confusion unless further commentary is provided.

Section 2.6 noted that linear feedback signals can create an amplification instability in networks whose function is to phasically store activity patterns in STM. In networks that react to tonically active input sources, such as the HMI or tonic cells, the amplification instability becomes the functionally useful property of temporally maintaining a stable baseline of activity or tone. A simple network that instantiates this concept is

$$\frac{d}{dt} S_i = -C S_i + (D - S_i)(I_i + S_i) - S_i \sum_{r\neq i} (I_r + S_r), \quad (6.22)$$

(Grossberg, 1973, 1978b), where the $I_i$ are the inputs to $F_1$ due to the PTS shifts and the $S_i$ are the normalized activations caused by these inputs.
within $F_1$. For simplicity, suppose that the total input $\sum_i I_i$ to (6.22) is small compared to the total number of sites $D$ in each population, and let $D > C$. Then, in response to any PTS shift pattern $(I_1, I_2, I_3, \ldots)$, the total activity $\sum_{i=1} S_i$ of (6.22) approximates the constant value $D - C$ and $S_i$ becomes proportional to $I_i$. When a new PTS shift pattern is input to the network, the total activity remains normalized as the normalized activities of $S_i$ adjust themselves to become proportional to the new inputs $I_i$.

These several spatial mapping models—which have used antagonistic positional gradients, contrast enhancement, normalization, coincidence detectors, PTS shifts, and adaptive coding mechanisms—define a domain of testable possibilities which future experiments need to address. Of particular interest is the question of whether an RM and an EPM are both designed using the same mechanisms, or whether the different prepro- cessing requirements of HMI outputs and T cell outputs have induced evolutionary specializations of their spatial mapping mechanisms.