Population Codes: Theory & Practice

Richard Zemel

Bruce McNaughton

NIPS00 Tutorial

November 27, 2000

Single cell: practice & theory

Single-cell analysis of neural coding took off in early 1960s: Hubel & Wiesel

clear techniques for:

- stimulating and recording from single cells
- varying conditions through stimulus, pharmacology
- lesioning out select groups of neurons

orientation tuning of V1 cells demonstrated correlate between stimulus properties and neural responses

also showed organization of visual cortex: hypercolumns or 'ice cube' model

Localist encoding

How to interpret this organization?

localist (identity) coding: each neuron well-tuned to single value of low-dimensional variable (e.g., orientation, spatial position, eye of input)

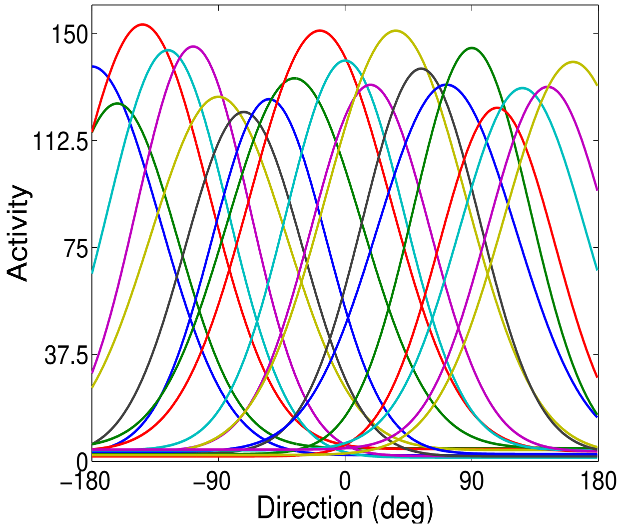
binary units (on or off)

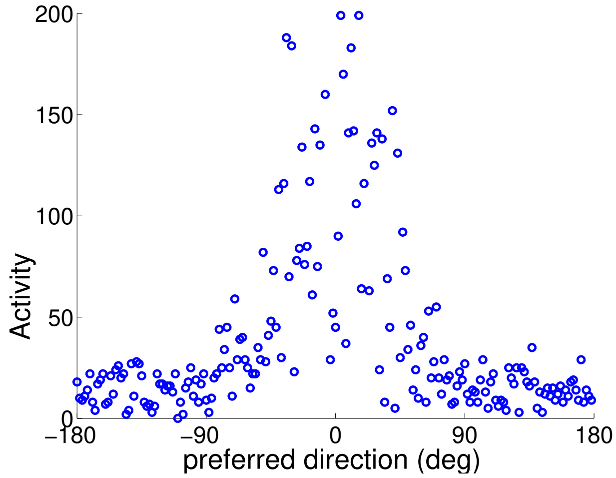
Barlow (1972) suggested that this form of representation applied throughout cortex, gaining in specificity as information ascended the processing hierarchy

1000 active neurons represent visual scene, each neuron represents equivalent of one word

Population codes

alternative idea: neural response more informative if actual value is relevant





population code: relative intensity of response of population of cells specifies value

in ideal case, only need two units/cells

but if noise present in responses – need larger set to specify value

key property of population codes: **robustness**

Early population code theory

population codes not entirely modern conception

color circle (Newton; Young); each color represented as point in 3-dimensional space, spanned by color primaries (RGV)

experiments on color mixtures (Maxwell, 1855) showed that any color can be represented as mathematical equivalent to mixture of 3 primaries

direct correspondence to sensory physiology – 3 types of cones

line-element models – form of population code interpretation:

- combination weights based on perceived discrimination between color stimuli (Helmholtz, 1896)
- can model interpretation as statistical discrimination of two color signals where underlying responses are noisy

Population codes all the way down

- visual system: spatial frequency, orienation, motion speed and direction, binocular disparity

- auditory: localization via interaural time and intensity differences

- bat echolocation: echo delay and Doppler shift

- olfactory, gustatory stimuli

- motor cortex: movement direction

- inferotemporal cortex: faces and complex objects

- somatosensory: body location

- hippocampus: spatial position

Outline

Today we focus on:

1. Simple methods of interpreting population codes
2. State-of-the-art in recordings from populations
3. Examples of population recordings in hippocampus
4. Optimal reconstruction methods
5. How can population encode uncertainty?
6. What computational power is provided by recurrent connections within a population?

Issue 1: Reading the neural code

Neurophysiologists collect neural recordings – sequences of action potentials (spikes) from one or several cells during controlled experiment

Task: reconstruct stimulus identity, or value of parameter(s) from responses

Why play the homunculus?

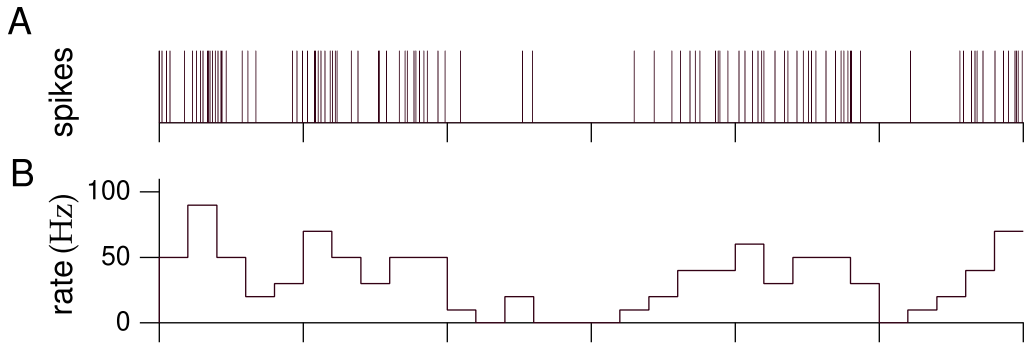
1. assess the degree to which that parameter encoded in population

2. determine limits on reliability and accuracy of neuronal encoding

3. characterize information processing – nervous system faced with this decoding problem

Rate representation of responses

action potentials convey information through timing



summarized in r_i: firing rate of cell i (number of spikes fired in particular time interval)

interval size determines amount of information about spike timing (A) lost in rate representation (B)

Example: Reconstructing movement direction

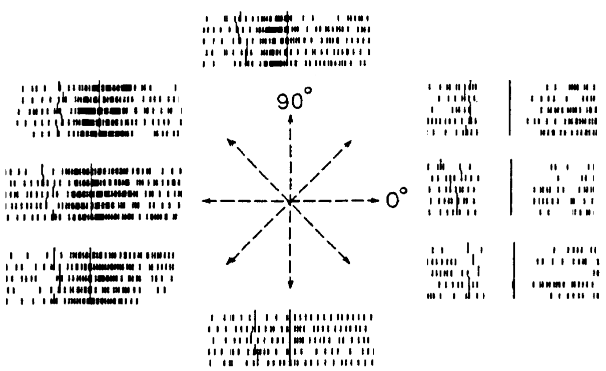
task: given rates $\mathbf{r} = r_1, \ldots, r_N$, compute arm direction

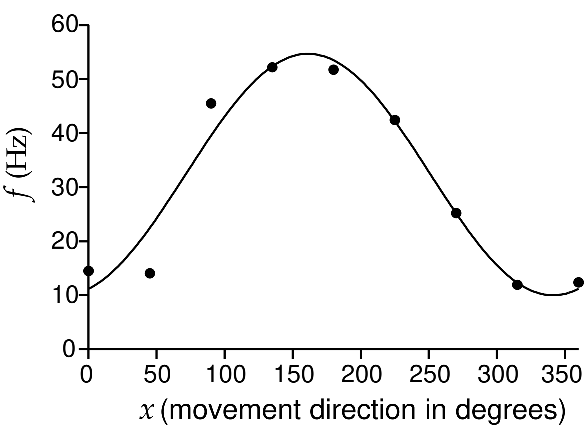
single cells in motor cortex (M1) tuned to movement angle (x), preferred angle x_i

tuning functions $f_i(x)$:

$$f_i(x) = A + B \cos(x - x_i)$$

$$A = \frac{1}{2}(r_{iMax} + r_{iMin}) \quad B = \frac{1}{2}(r_{iMax} - r_{iMin})$$

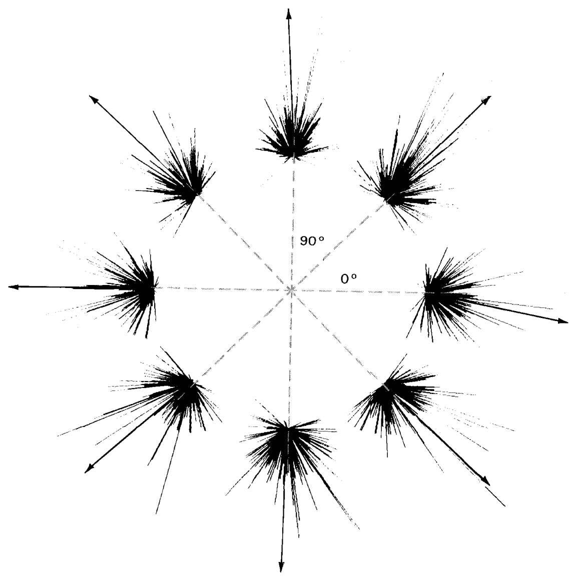




Population vector method

- each cell represented as vector pointing in preferred direction $\vec{x}_i$
- length of vector represents relative response strength for particular movement (can scale r_i by subtracting off B)
- sum of vectors describes estimated movement direction $\equiv$ the *population vector*

$$\vec{x}_{pop} = \sum_i r_i \vec{x}_i$$



simple, robust, accurate method if N large, and $\{\vec{x}_i\}$ randomly, uniformly span space of directions

Reconstruction via Basis Functions

population vector method yields single vector $\vec{x}_{pop}$

a *basis function method* yields scalar function over $\vec{x}$:

$$\rho(x) = \sum_{i=1}^{N} r_i \psi_i(x)$$

where $\psi_i(x)$: kernel, or basis function

can extract peak value as reconstructed value of x:

$$x_{basis} = \arg\max_x \rho(x)$$

basis can be pre-defined, e.g.,

- **template method**: tuning functions as basis set

$$x_{template} = \arg\max_x \sum_i r_i f_i(x)$$

- **population vector**: cosine basis $\psi_i(x) = \vec{x}_i \cdot \vec{x}$

$$\arg\max_{\vec{x}} \sum_i r_i \vec{x}_i \cdot \vec{x} \text{ matches } \vec{x}_{pop} \text{ direction}$$

Optimized Basis Functions

instead of pre-defining basis set, can use numerical methods to determine optimal basis set $\{\psi_i(x)\}$

for example, least-squares regression: know true distribution $\rho_*^t(x)$ in instance t, minimize

$$E = \sum_{t,x} \left[\rho_*^t(x) - \sum_i r_i^t \psi_i(x) \right]^2$$

general basis function methods require more information than population vector method, but some advantages obtained:

- less sensitive to noise
- can take into account non-uniform $f_i(x)$, preferred directions x_i

4. Bayesian reconstruction

linear methods are simple and can perform well, but other class of methods are in some sense optimal

begin with *encoding* model:

$$r_i(x) = f_i(x) + \eta$$

- rate $r_i(\mathbf{x})$ as random variable – response of cell i in population to stimulus value x
- tuning function $f_i(\mathbf{x})$ describes expected rate
- noise η typically assumed Poisson or Gaussian

goal: *decode* responses $\mathbf{r}$ to form posterior distribution

$$P(x|\mathbf{r}) = \frac{P(\mathbf{r}|x)P(x)}{P(\mathbf{r})}$$

- $P(x)$ based on distribution of stimulus values
- $P(\mathbf{r})$ is prior over rates, can be based on tuning functions if available: $P(\mathbf{r}) = \int_x P(\mathbf{r}|x)P(x)$

Standard Bayesian reconstruction

key - specifying likelihood $P(\mathbf{r}|x)$

assumptions:

- spikes have Poisson distribution (counting spikes in small window when distributed independently and randomly)
- noise uncorrelated between different cells (all variability captured in $P(r_i|x)$)

$$P(\mathbf{r}|x) = \prod_{i=1}^{N} P(r_i|x) = k(\mathbf{r}) \prod_{i=1}^{N} f_i(x)^{r_i} \exp(-f_i(x))$$

intuition – gain higher precision through multiplying rather than adding tuning curves

obtain probability distribution, extract single value:

1. MAP: $x_{MAP} = \arg\max_x P(x|\mathbf{r})$
2. Maximum Likelihood (ML): $x_{ML} = \arg\max_x P(\mathbf{r}|x)$

Application: hippocampal population codes

- $P(x)$ based on spatial occupancy
- $P(r_i|x)$ are *place fields*

reconstruction improved considerably by using continuity constraint: $P(x^t|x^{t-1})$ peaked around x^{t-1}

More on ML

can implement ML by maximizing

$$\ln P(\mathbf{r}|x) = \sum_{i=1}^{N} r_i \ln f_i(x) + f_i(x)$$

for evenly distributed tuning curves ($\sum_i f_i(x)$ constant), this implies that

$$\sum_{i=1}^{N} r_i \frac{f_i'(x)}{f_i(x)} = 0$$

where $f_i'(x)$ is the derivative of the tuning curve

for Gaussian tuning curves, $\frac{f_i'(x)}{f_i(x)} = (x_i - x)/\sigma^2$:

$$x_{ML} = \sum_i \frac{r_i x_i}{r_i}$$

Optimality of Bayesian Method

expected squared error of decoding is sum of

- squared **bias** [$b^2(x)$: decoder mean vs. true value]
- **variance** [$\sigma^2(x)$: decoder output vs. its mean]

Cramer-Rao bound places limits on accuracy of any decoding method:

$$\sigma^2(x) \geq \frac{(1 + b'(x))^2}{I_F(x)}$$

where $I_F(x)$ is the *Fisher information*:

$$I_F(x) = \left\langle -\frac{\partial^2 \ln P(\mathbf{r}|x)}{\partial x^2} \right\rangle$$

if estimation unbiased:

$$\sigma^2(x) \geq \frac{1}{I_F(x)}$$

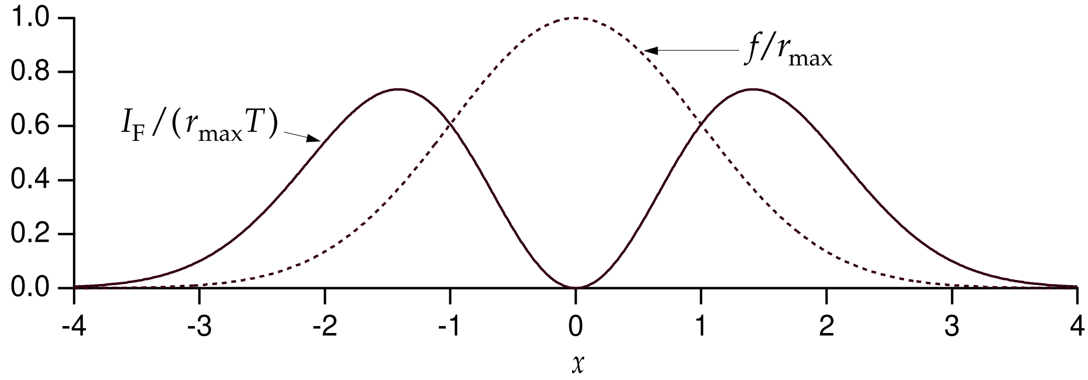
Bayesian decoding (with uniform prior) equivalent to maximum likelihood estimation, which reaches Cramer-Rao bound in large sample limit

Determining relevant units

can use same analysis to determine which units contribute most to decoding

for symmetric tuning curves:

$$I_F(x) = T \sum_{i=1}^{N} \frac{(f_i'(x))^2}{f_i(x)}$$



most important units:

- not those whose tuning curve peak at estimate
- instead those for which estimate lies on part of tuning curve with highest derivative

5. Extending information in population codes

Standard model aims to estimate single value of x from population responses $\mathbf{r}$

Need to extend this model to handle some important situations

main claim:

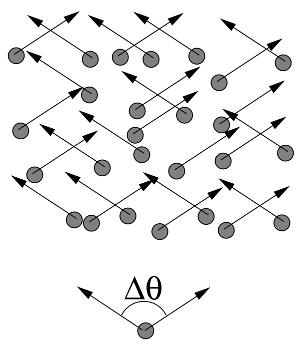
populations represent more than single value

Two encoding problems:

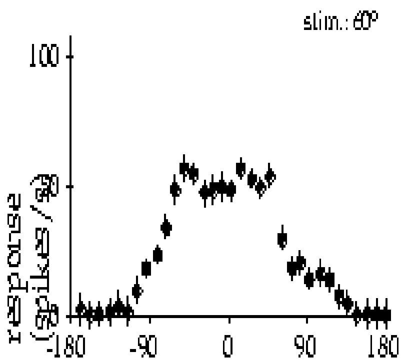
(1). *multiple values*

e.g., motion transparency: several directions of motion perceived at same location

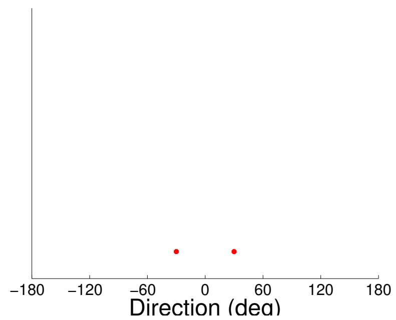
stimulus



MT response



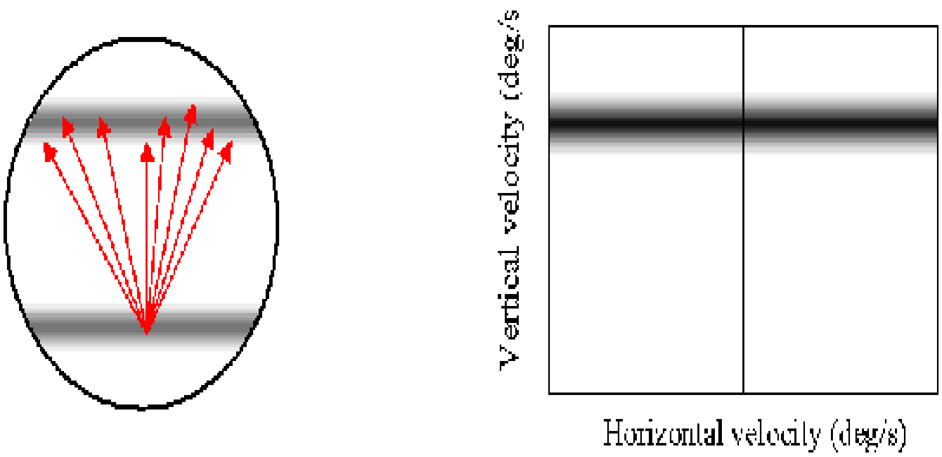
subject's response



Encoding uncertainty in population codes

(2). *uncertainty*: noise (at all levels); inherent in image (insuffcent information)

e.g., aperture problem: cannot extract component of motion in direction along which stimulus constant



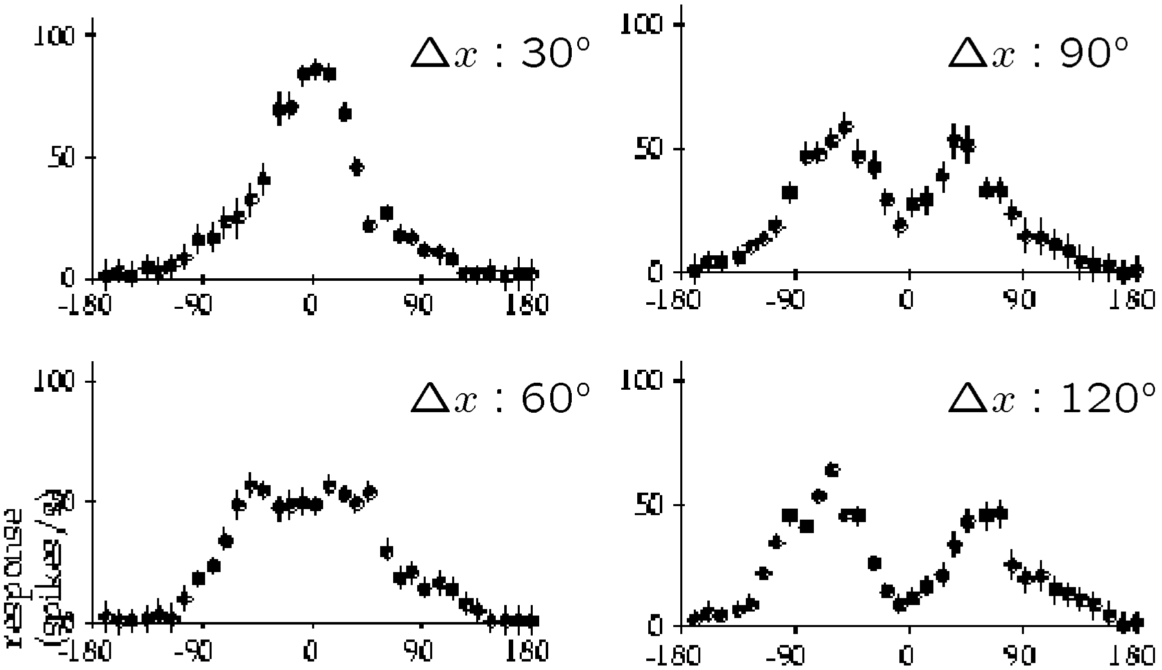
uncertainty common: low contrast images, noisy data

can represent uncertainty by exciting all units in population whose preferred value is consistent with input

same representation can be given to visual input containing multiple values

but can standard model handle this?

Standard Bayesian Decoding Fails



$$\log P[x|\,\{r_i\}] \sim \log P[x] \prod_i P[r_i|x] \sim \sum_i r_i \log f_i(x)$$

under standard Poisson, conditional independence assumptions

if tuning functions are unimodal, decoding convolves responses with unimodal kernels: *unimodal* response pattern produces unimodal $\log P[x|\,\{r_i\}]$; also fails on *bimodal* response patterns

$\Rightarrow$ only extracts one motion component from responses to transparent motion

Distributional Population Coding

Extend standard encoding model to allow $\mathbf{r}$ to depend on general $P[x]$:

$$r_i = \int_x g[x] f_i(x) dx + \eta$$

where $g[x]$ specifies what motions are consistent with image

decoding methods:

1. basis functions

$$g[x] \sim \sum_i r_i \phi_i(x)$$

2. approximate ML in distributions over x

$$g[x] = P[x|\mathbf{r}]$$

$$\log P[x|\mathbf{r}] \sim \sum_i r_i \log\left[f_i(x) * P(x|\mathbf{r})\right] - \alpha h\left[P(x|\mathbf{r})\right]$$

where $h()$ is a regularizer (simple smoothness prior)

6. Processing in recurrent connections

Most of computational focus on population codes based on observation that they offer balance between localist and fully distributed encoding

- Localist encoding: problems with noise; robustness; number of cells required
- Fully distributed coding: decoding complicated; cannot handle multiple values

Other computational properties of population codes studied recently

Key focus (driven partly by biological studies): recurrent connections between units in population

Line Attractors

Simple network model, with recurrent connections T_{ij} governed by dynamic equation

$$\tau \frac{\mathrm{d}u_i}{\mathrm{d}t} + u_i = \sum_{j=1}^{N} T_{ij} r_j + h_i$$

where u_i is input into unit i; rate r_i is its output; h_i is feedforward contribution; $\sum_{j=1}^{N} T_{ij} r_j$ recurrent

if rate linear above threshold input: $r_i = g_i u_i + r_{iBase}$, fixed-point equations:

$$r_i = \sum_{j=1}^{N} W_{ij} r_j + f_i$$

- $f_i = g_i h_i + r_{iBase}$ is feedforward contributions
- $W_{ij} = g_i T_{ij}$: recurrent contribution of j on i

in general this set of N linear equations in N unknowns r_i has unique solution

but with appropriate tuning of connections (e.g., symmetric weight matrix, with single unity eigenvalue), fixed points will lie along a line

Implementing short-term memory

This formulation applied to a number of problems

First consider short-term memory: try to remember direction you are facing in a room after closing your eyes

head direction cells in rats: smooth bell-shaped tuning functions to direction of head in room; maintain discharge even in dark

can construct simple network model:

- symmetric recurrent connections in population (local excitation, lateral inhibition)
- nonlinear activation function

peak position of hill of activity memory for value stable states

problems with stability (drift over time); updating

Noise removal

Same circuit may be used to clean up noisy population responses

with transient noisy input, network settles into hill of activity

- shape of hill corresponds to tuning functions
- peak position is close approximation to x_{ML}

smooth hill centered on any point is a stable activity profile: *line attractor*

this noise removal process in the population allows simple decoding techniques (such as population vector method) to approximate ML

Selection and gain modulation

Recurrent connections in population can also produce two different forms of behavior:

1. stimulus selection (nonlinear): strongest of multiple signals at different locations selected, weaker signals suppressed

2. gain modulation (linear): population response levels multiplicatively scaled by background amplitude

This can be achieved through the same type of recurrent connections:

- strong positive recurrent connections select active neurons

- active neurons generate weaker feedback that amplifies input in analog manner

This type of circuit recently implemented in silicon

Analysis shows that for symmetric connections $W_{ij} = W_{ji}$, the set of stable states can be defined

Dynamic synapses

Recent twist to the computation via recurrent connections story: weights of variable strength

biological inspiration: fast synaptic dynamics

one scheme suggests that strength of recurrent excitation decreases after stimulus onset: shift from nonlinear, competitive regime to linear more cooperative phase

this allows the shift from nonlinear to linear processing to occur in time

Selected References

Extracting information from population codes

Foldiak (1993). In *Computation and Neural Systems*. Eeckman & Bower (Eds.), Kluwer Academic Publishers: 55-60.

Georgopoulos, A. et al. (1982). *Journal of Neuroscience* 2(11): 1527-1537.

Salinas, E. & L. Abbott (1994). *Journal of Computational Neuroscience* 1: 89-108.

Zhang, K., et al. (1998). *Journal of Neurophysiology* 79(2): 1017-44.

Representing uncertainty in population codes

Anderson, C. (1994). In *Computational Intelligence Imitating Life*. IEEE Press: 213-222.

Sanger, T. (1996). *Journal of Neurophysiology* 76(4): 2790-3.

Simoncelli, E., et al. (1991). In *Proceedings 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*: 310-315.

Treue, S., et al. (2000). In *Nature Neuroscience* 3(3): 270-6.

Zemel, R. & P. Dayan (1999). In *Advances in Neural Information Processing Systems 11*. Kearns, Solla, & Cohn (Eds). MIT Press: 174-180.

Zemel, R., et al. (1998). *Neural Computation* 10: 403-430.

Recurrent computation in population codes

Adorjan, P, et al. (2000). In *Advances in Neural Information Processing Systems 12*. Solla, Leen, & Muller (Eds). MIT Press: 89-96.

Deneve, S., et al. (1999). *Nature Neuroscience* 2(8): 740-745.

Hahnloser, R., et al. (2000). *Nature* 405: 947-951.

Salinas, E. & Abbott, L. (1997). *J Neurophysiol* 77(6): 3267-72.

Seung, H. (1996). *Proc Natl Acad Sci USA* 93(23): 13339-44.

Zhang, K. (1996). *Journal of Neuroscience* 16(6): 2112-2126.