IT neurons are tolerant to identity-preserving transformations

Rust & DiCarlo, 2012
The geometry of selectivity and invariance. The three axes are three image dimensions (e.g., the values of three pixels in an image). Real images require several thousand dimensions, but we use three for simple visualization. Any point in the space corresponds to a different image. The gray surface represents a continuous subset, or manifold, of images of a particular object. If a hypothetical neural population effectively encodes this object's identity, all object images from this manifold will yield patterns of neural responses that are distinguishable from the patterns of responses induced by other sets of images. Moving along the surface of the manifold changes the image itself but maintains the ability of the neural population to discriminate the image from others. This is a direction of invariance. Moving away from, or orthogonal to, the surface of the manifold changes the image in a way that prevents the population from effectively discriminating. This is a direction of selectivity. The manifold shown here corresponds to a set of population responses that are selective for proboscis monkeys, not just for image patches with similar color and texture, but are also invariant to changes in size (near vs far) and context (face only vs face and body).

Freeman & Ziemba, 2011
Object tangling

DiCarlo & Cox, 2007
Untangling object manifolds along the ventral visual stream

(a) V1 space

(b) IT space

(c) ‘Classical’ IT space

Response vs. Pose

Response vs. Pose

Response vs. Pose

TRENDS in Cognitive Sciences

DiCarlo & Cox, 2007
The form processing pathway maintains an “equally distributed” representation of images.
The neural activity was further analyzed by constructing average spike density functions (SDFs), sorted by the monkey's perceptual reports. Fig. 4 shows these data for the same cell depicted in the Fig. 3. Fig. 4A and B show responses in nonrivalrous and rivalrous conditions, respectively. As shown in Fig. 3A, this neuron fired vigorously when the monkey reported seeing the cell's preferred pattern in both the nonrivalrous and rivalrous conditions. However, when the monkey reported seeing the ineffective stimulus, the cell response was almost eliminated, even when the effective stimulus was physically present during rivalry.

To increase the instances of exclusive visibility of one stimulus, and to further ensure that the monkey's report accurately reflected which stimulus he perceived at any given time, we also tested the psychophysical performance of the monkeys and the neural responses of STS and IT cells using the flash suppression paradigm (10). In this condition, one of the two stimuli used to instigate rivalry is first viewed monocularly for 1–2 sec. Following the monocular preview, rivalry is induced by presenting the second image to the contralateral eye. Under these conditions, human subjects invariably perceive only the newly presented image and the previewed stimulus is rendered invisible. Previous studies have shown that the suppression of the previewed stimulus is not due to forward masking or light adaptation (10) and that instead it shares much in common with the perceptual suppression experienced during binocular rivalry (11). In our experiments, the monkeys, just like the human subjects,
consistently reported seeing the stimulus presented to the eye contralateral to the previewing eye during the flash suppression trials.

To confirm that the animals responded only when a flashed stimulus was exclusively dominant, catch trials were introduced in which mixed stimuli were flashed, after which the monkey was required to release both levers. Performance for both animals was consistently 95% for this task. Fig. 4B shows the activity of an STS neuron in the flash suppression condition. Fig. 4B upper shows the cell responses for monocular presentations, and the Fig. 4B lower shows the neuron's activity at the end of the monocular preview (to the left of the dotted vertical line) and when perceptual dominance is exogenously reversed as the rival stimulus is presented to the other eye (to the right of dotted vertical line). The cell fires vigorously when the effective stimulus dominates perception and ceases firing entirely when the ineffective stimulus is made dominant. To better understand the differences between the temporal areas and the prestriate areas, recordings were also performed in area V4 using the flash suppression paradigm (D. Leopold and N.K.L., unpublished observations). V4 neurons were largely unaffected by the perceptual changes during flash suppression. Presenting the ineffective stimulus after priming with the effective one caused no alteration in the firing rate of any of the neurons.
Correlation of IT activity and perceptual state during binocular rivalry (Logothetis, 1998)

1810 N. K. Logothetis Single units and conscious vision

[Diagram showing brain regions V1, V2, V4, MT (V5), TPO, TEm, TEa with corresponding frequency data for excited and suppressed states.]
Ungerleider & Mishkin, 1982

**Ventral pathway**
Form, recognition, memory

**Dorsal pathway**
Space, motion, action

Ungerleider & Mishkin, 1982
Why motion?

George Mather, Patrick Cavanagh, and others
First demonstration of direction selectivity in macaque MT/V5 by Dubner & Zeki (1971). (a) Neuronal responses to a bar of light swept across the receptive field in different directions (modified from figure 1 of Dubner & Zeki 1971). Each trace shows the spiking activity of the neuron as the bar was swept in the direction indicated by the arrow. The neuron’s preferred direction was up and to the right. (b) Oblique penetration through MT (modified from figure 3 of Dubner & Zeki 1971) showing the shifts in preferred direction indicative of the direction columns subsequently demonstrated by Albright et al. (1984). See also Figure 4.
 Movshon & Newsome, 1996

Proportion of cells

Directionality index

V1

MT
Movshon & Newsome, 1996
Center-surround interactions in MT. (A) Effect of contrast on center-surround interactions for one MT neuron. When tested with high-contrast random dots (RMS contrast 9.8 cd/m²) the neuron responded optimally to a circular dot patch 10° in diameter and was strongly suppressed by larger patterns. The same test using a low-contrast dot pattern (0.7 cd/m²) revealed strong area summation with increasing size. (B) Population of 110 MT neurons showing the strength of surround suppression measured at both high and low contrast. Surround suppression was quantified as the percent reduction in response between the largest dot patch (35° diameter) and the stimulus eliciting the maximal response. Each dot represents data from one neuron; the dashed diagonal is the locus of points for which the surround suppression was unchanged by contrast. The circled dot is the cell from panel A. (C) Asymmetries in the spatial organization of the suppressive surround (after Xiao et al. 1997). Different kinds of surround geometry are potentially useful for calculating spatial changes in flow fields that may be involved in the computation of structure from motion. Neurons whose receptive fields have circularly symmetric surrounds (top) are postulated to underlie figure-ground segregation. The first- (middle) and second-order (bottom) directional derivatives can be used to determine surface tilt (or slant) and surface curvature, respectively (Buracas & Albright 1996). Panels A and B are from Pack et al. 2005.
FIG. 5. Responses of a representative unit in MT to stimuli moving in its preferred direction at different speeds. In this and all subsequent plots the speed axis is logarithmic. Bars indicate the standard errors of the mean for five repetitions of each speed. A dashed line marks the background rate of firing. This unit, like most in MT, had a sharp peak in its response curve. Summed response histograms in the lower half of the figure show that the peak rate of firing closely follows the average rate of firing. Tic marks under each histogram denote times of stimulus onset and offset. The receptive field was 15° across and each stimulus traversed 20°.
Gratings, plaids, and coherent motion
Grating responses  

Plaid responses

V1 cell

MT component cell

MT pattern cell

Movshon, Adelson, Gizzi & Newsome, 1985
CORRELATION COMPONENT

CORRELATION PATTERN

Movshon & Newsome, 1996
MST also contains a high proportion of pattern cells.
Local field potentials may reveal stages in pattern computation
Local field potentials may reveal stages in pattern computation

Khawaja, Tsui & Pack, 2009
MT pattern cell

Components of the optimal plaid

Grating responses

Plaids containing the optimal grating

Plaid responses

Movshon, Adelson, Gizzi & Newsome, 1985
In search of a simple model

Lateral geniculate cells

Simple cortical cell

Movshon et al, 1985

Hubel & Wiesel, 1962

Simoncelli & Heeger, 1998
A simple and (mostly) feedforward model

Retinal image

Moving image

Linear operator
Gain control
Output nonlinearity

Linear operator
Gain control
Output nonlinearity

Simoncelli & Heeger., 1998
1D motion stimuli: gratings
2D motion stimuli: plaids
2D motion stimuli: textures
1D motion stimuli
Is pattern motion computed globally?

Majaj, Carandini & Movshon, 2007
Majaj, Carandini & Movshon, 2007
Small plaid

Pseudoplaids

Majaj, Carandini & Movshon, 2007
Pattern motion is computed locally
How do local and global motion signals interact?

Hedges, Gartshteyn, Kohn, Rust, Shadlen, Newsome & Movshon, 2011
How do local and global motion signals interact?

- **Global preferred**
  - Local preferred
- **Global preferred**
  - Local null
- **Global null**
  - Local preferred
- **Global null**
  - Local null

Hedges, Gartshteyn, Kohn, Rust, Shadlen, Newsome & Movshon, 2011
How do local and global motion signals interact?

Hedges, Gartshteyn, Kohn, Rust, Shadlen, Newsome & Movshon, 2011
How do local and global motion signals interact?

Hedges et al, 2011
How do local and global motion signals interact?

Proportion of cells

Local dominance

How do local and global motion signals interact?

Hedges, Gartshteyn, Kohn, Rust, Shadlen, Newsome & Movshon, 2011
A simple and (mostly) feedforward model

Retinal image

Linear operator

Gain control

Output nonlinearity

V1

Moving image

Linear operator

Gain control

Output nonlinearity

MT

Simoncelli & Heeger, 1998; Rust, Mante, Simoncelli & Movshon, 2006
Component cell

Pattern cell

Rust, Mante, Simoncelli & Movshon, 2006
Direction-interaction: 
Gratings
Direction-interaction:

Plaids
Direction-interaction:
One common component
Direction-interaction:
Common axis
Component cell

Pattern cell

Rust, Mante, Simoncelli & Movshon, 2006
Recovered model elements

**Data**

- Direction
- 360
- 180
- 0

**Model**

- V1 gain control
- Normalization strength
- Connection weights
- MT linear weights
- Preferred direction of V1 input neuron
Pattern direction selectivity arises from:

1. Broad convergence of excitatory inputs
2. Strong motion opponent suppression
3. Strong tuned gain control

Rust, Mante, Simoncelli & Movshon, 2006
Limitations of the approach

MT receptive field
Spatial and spectral structure of motion-enhanced natural movies

Nishimoto & Gallant, 2011
“Motion-enhanced” natural movies

Nishimoto & Gallant, 2011
“Motion-enhanced” natural movies, and friends
Analysis of MT neurons using a “boosted” model

Nishimoto & Gallant, 2011
Estimated spectral receptive fields of four MT neurons

Nishimoto & Gallant, 2011
MT neurons vary in the degree to which their excitatory spectral receptive fields form a ring within the optimal velocity plane.

Nishimoto & Gallant, 2011
Two neural correlates of consciousness

Ned Block

Block’s conjecture

MT is “the core phenomenal neural correlate of consciousness for the visual experiential content as of motion”
Local and global motion signals