Conventional Wisdom About the Functional Implications of the Anatomical Hierarchy

It would appear rather straightforward to assume that the functional organization of the visual system somehow directly reflects the underlying anatomical hierarchy. In its extreme form, there are two facets to this assumption. One is that the visual processing itself is hierarchical, and the other is that the hierarchy at the functional level parallels that at the anatomical level.

The notion that visual processing is hierarchical has been around since before the anatomical hierarchy was elucidated. Marr (1982) was one of the early and influential proponents of hierarchical processing in vision (Fig. 1). He proposed that during the early stages of visual processing, the visual system extracts information about the local image elements (i.e., the basic “building blocks” or primitives) of the visual scene, such as the local contrast, orientation, and so on, to construct a raw “primal sketch” of the visual scene. In intermediate stages of processing, the visual system constructs a representation of object surfaces (or “2½-D sketch”) using the information about the primitives that was extracted during the previous stage. Finally, the visual system constructs a full representation of the visual scene (or “3-D sketch”) by combining the various elements of the 2½-D sketch. Many modern models also propose similar processing hierarchies (see Palmer 1999).

Fig. 1. Anatomical and functional hierarchies in the macaque visual system. The human visual system (not shown) is believed to be roughly similar.

A, A schematic summary of the laminar patterns of feed-forward (or ascending) and feed-back (or descending) connections for visual area V1. The laminar patterns vary somewhat from one visual area to the next. But in general, the connections are complementary, so that the ascending connections terminate in the granular layer (layer 4) and the descending connections avoid it. The connections are generally reciprocal, in that an area that sends feed-forward connections to another area also receives feedback connections from it. The visual anatomical hierarchy is defined based on, among other things, the laminar patterns of these interconnections among the various areas. See text for details.

B, A simplified version of the visual anatomical hierarchy in the macaque monkey. For the complete version, see Felleman and Van Essen (1991). See text for additional details. AIT = anterior inferotemporal; LGN = lateral geniculate nucleus; LIP = lateral intraparietal; MT = middle temporal; MST = medial superior temporal; PIT = posterior inferotemporal; V1 = visual area 1; V2 = visual area 2; V4 = visual area 4; VIP = ventral intraparietal.

C, A model of hierarchical processing of visual information proposed by David Marr (1982).

D, A schematic illustration of the presumed parallels between the anatomical and functional hierarchies. It is widely presumed not only that visual processing is hierarchical but also that the anatomical hierarchy provides a substrate for, and therefore parallels, the hierarchical processing.

Table 1. Connectivity of Areas/Regions in a Hypothetical Visual System

<table>
<thead>
<tr>
<th>Receives</th>
<th>Sends</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cortical Area/ Ascending</td>
<td>Ascending</td>
</tr>
<tr>
<td>Subcortical Nucleus Input from Output to</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>B, D</td>
</tr>
<tr>
<td>B</td>
<td>A, F</td>
</tr>
<tr>
<td>C</td>
<td>D, G</td>
</tr>
<tr>
<td>D</td>
<td>C, E, F</td>
</tr>
<tr>
<td>E</td>
<td>D, F</td>
</tr>
<tr>
<td>F</td>
<td>D, E, H</td>
</tr>
<tr>
<td>G</td>
<td>H, I</td>
</tr>
<tr>
<td>H</td>
<td>G, I</td>
</tr>
<tr>
<td>I</td>
<td>J</td>
</tr>
<tr>
<td>J</td>
<td>K</td>
</tr>
<tr>
<td>K</td>
<td>Retina</td>
</tr>
</tbody>
</table>

Only the ascending connections are shown. Given these connectivity data, can you arrange these areas into a hierarchy? The answer is shown below. Does the hierarchy remain the same if the input source for K is unknown? What happens if other inputs and/or outputs are unknown? (Answer: This data set results in the same hierarchical structure shown in Figure 1B, with the names of the visual areas/regions substituted as follows: A = 7a; B = AIT; C = VIP; D = MST; E = LIP; F = PIT; G = MT; H = V4; I = V2; J = V1; K = LGN.)
VI simple cells

Hubel and Wiesel, 1959

DeAngelis, et al, 1995
Out of the retina

- > 23 distinct neural pathways, no simple function division
- *Suprachiasmatic nucleus*: circadian rhythm
- *Accessory optic system*: stabilize retinal image
- *Superior colliculus*: integrate visual and auditory information with head movements, direct eyes
- *Pretectum*: plays adjusting pupil size, track large moving objects
- *Pregeniculate*: cells responsive to ambient light
- *Lateral geniculate (LGN)*: main “relay” to visual cortex; contains 6 distinct layers, each with 2 sublayers. Organization very complex
VI simple cell integration of LGN cells

Visual field

Brain

LGN cells

Simple cell

Hubel and Wiesel, 1963
Hyper-column model

Hubel and Wiesel, 1978
Anatomical circuitry in V1

From Callaway, 1998
Where is this headed?
A general approach to coding: redundancy reduction

Redundancy reduction is equivalent to efficient coding.
Reducing pixel redundancy

Lena: a standard 8 bit 256x256 gray scale image

histogram of pixel values
Entropy = 7.57 bits

from Daugman (1990)
2D Gabor wavelet captures spatial structure

- 2D Gabor functions

- Wavelet basis generated by dilations, translations, and rotations of a single basis function

- Can also control phase and aspect ratio

- (drifting) Gabor functions are what the eye “sees best”
Recoding with Gabor functions

Pixel entropy = 7.57 bits

Recoding with 2D Gabor functions
Coefficient entropy = 2.55 bits
Describing signals with a simple statistical model

**Principle**

*Good codes capture the statistical distribution of sensory patterns.*

How do we describe the distribution?

- Goal is to encode the data to desired precision

\[
x = \vec{a}_1 s_1 + \vec{a}_2 s_2 + \cdots + \vec{a}_L s_L + \vec{\epsilon}
\]

\[
= As + \epsilon
\]

- Can solve for the coefficients in the no noise case

\[
\hat{s} = A^{-1}x
\]
An algorithm for deriving efficient linear codes: ICA

Learning objective:

\( \text{maximize coding efficiency} \)

\( \Rightarrow \text{maximize } P(x|A) \text{ over } A. \)

Probability of the pattern ensemble is:

\[ P(x_1, x_2, \ldots, x_N|A) = \prod_{k} P(x_k|A) \]

To obtain \( P(x|A) \) marginalize over \( s \):

\[ P(x|A) = \int ds \, P(x|A, s)P(s) \]

\[ = \frac{P(s)}{|\det A|} \]

Using independent component analysis (ICA) to optimize \( A \):

\[
\Delta A \propto AA^T \frac{\partial}{\partial A} \log P(x|A) \\
= -A(zs^T - I)
\]

where \( z = (\log P(s))' \).

This learning rule:

- learns the features that capture the most structure
- optimizes the efficiency of the code

What should we use for \( P(s) \)?
Modeling Non-Gaussian distributions

- Typical coeff. distributions of natural signals are *non-Gaussian*.
The generalized Gaussian distribution

\[ P(x|q) \propto \exp\left(-\frac{1}{2}|x|^q\right) \]

- Or equivalently, and exponential power distribution (Box and Tiao, 1973):

\[ P(x|\mu, \sigma, \beta) = \frac{\omega(\beta)}{\sigma} \exp\left[-c(\beta)\left|\frac{x-\mu}{\sigma}\right|^{2/(1+\beta)}\right] \]

- \( \beta \) varies monotonically with the kurtosis, \( \gamma_2 \):

\[
\begin{align*}
\beta & = -0.75 \quad \gamma_2 = -1.08 \\
\beta & = -0.25 \quad \gamma_2 = -0.45 \\
\beta & = +0.00 \quad \gamma_2 = +0.00 \\
\beta & = +0.50 \quad \gamma_2 = +1.21 \\
\beta & = +1.00 \quad \gamma_2 = +3.00 \\
\beta & = +2.00 \quad \gamma_2 = +9.26
\end{align*}
\]
Modeling Gaussian distributions with PCA

- Principal component analysis (PCA) describes the principal axes of variation in the data distribution.
- This is equivalent to fitting the data with a multivariate Gaussian.
Modeling non-Gaussian distributions

- What about non-Gaussian marginals?

- How would this distribution be modeled by PCA?
Modeling non-Gaussian distributions

- What about non-Gaussian marginals?

- How would this distribution be modeled by PCA?

- How should the distribution be described?

The non-orthogonal ICA solution captures the non-Gaussian structure
Efficient coding of natural images

Network weights are adapted to maximize coding efficiency:
minimizes redundancy and maximizes the independence of the outputs
Model predicts local and global receptive field properties

Learned basis for natural images

Overlaid basis function properties

from Lewicki and Olshausen, 1999
Algorithm selects best of many possible sensory codes

Learned

Wavelet

Haar

Gabor

Fourier

PCA

Theoretical perspective: Not edge “detectors.” An efficient code for natural images.

from Lewicki and Olshausen, 1999
2D Receptive fields in primary visual cortex

2D Receptive Field

2D Gabor Function

Difference

Fit of 2D Gabor wavelet is indistinguishable from noise.

figure from Daugman, 1990
data from Jones and Palmer, 1987
Comparing coding efficiency on natural images

![Comparison of coding efficiency on natural images](image)

- Gabor
- Wavelet
- Haar
- Fourier
- PCA
- Learned

Estimated bits per pixel
Comparing efficiency estimates using entropy and probability

Entropy estimate ignores fidelity

Learned
GaborFit
ICA
PCA
Fourier
Haar
Pixel

Estimated bits per pixel

- estimate based on $P(x|A)$
- entropy estimate
Optimality depends on data

Which code will be best for random, sparse pixels?
Now the coding efficiencies are reversed

The pixel basis is now optimal.

<table>
<thead>
<tr>
<th>Basis</th>
<th>Estimated Bits per Pixel</th>
</tr>
</thead>
<tbody>
<tr>
<td>learned</td>
<td>7.5 ± 0.2</td>
</tr>
<tr>
<td>GaborFit</td>
<td>9.0 ± 0.3</td>
</tr>
<tr>
<td>ICA</td>
<td>6.0 ± 0.1</td>
</tr>
<tr>
<td>PCA</td>
<td>5.0 ± 0.1</td>
</tr>
<tr>
<td>Fourier</td>
<td>5.0 ± 0.1</td>
</tr>
<tr>
<td>Haar</td>
<td>5.0 ± 0.1</td>
</tr>
<tr>
<td>pixel</td>
<td>5.0 ± 0.1</td>
</tr>
</tbody>
</table>

- **White bars**: Estimate based on $P(x|A)$
- **Black bars**: Entropy estimate
Responses in primary visual cortex to visual motion

from Wandell, 1995
Sparse coding of time-varying images (Olshausen, 2002)

\[ I(x, y, t) = \sum_i \sum_{t'} a_i(t') \phi_i(x, y, t - t') + \epsilon(x, y, t) \]

\[ = \sum_i a_i(t) * \phi_i(x, y, t) + \epsilon(x, y, t) \]
Sparse decomposition of image sequences

**convolution**

**posterior maximum**

reconstruction

input sequence

from Olshausen, 2002
Learned spatio-temporal basis functions

from Olshausen, 2002
Animated spatial-temporal basis functions

from Olshausen, 2002
Redundancy reduction for noisy channels (Atick, 1992)

\[ y = Ax + \delta \]

Mutual information

\[ I(x, s) = \sum_{s,x} P(x, s) \log_2 \left[ \frac{P(x, s)}{P(s)P(x)} \right] \]

\[ I(x, s) = 0 \text{ iff } P(x, s) = P(x)P(s), \text{ i.e. } x \text{ and } s \text{ are independent.} \]
A second order statistical model: Gaussian

- To calculate $I(x,s)$, we need $P(s)$, $P(x)$, and $P(x,s)$
- Assume it is sufficient to measure 2nd order correlations
  (this is equivalent to measuring the avg. spatial frequency):

\[
\begin{align*}
\langle s[n]s[m]\rangle &= R_0[n,m] \\
\langle x[n]x[m]\rangle &= R_0[n,m] + N^2\delta_{n,m} \equiv R[n,m] \\
\langle x[n]s[m]\rangle &= \langle s[n]s[m]\rangle \\
\end{align*}
\]

for $u = s, x, y$ and $R_{uu}[n,m] \equiv \langle u[n]u[m]\rangle$.

\[
P(u) = [(2\pi)^d \det(R_{uu})]^{-1/2} \exp\left[-\frac{1}{2} \sum_{n,m} (u[n] - \bar{u}) R^{-1}_{uu}[n,m](u[m] - \bar{u})\right]
\]

\[
= \mathcal{N}(\bar{u}, R_{uu})
\]
Mutual information between stages of the model

\[
I(x, s) = \frac{1}{2} \left[ \frac{\det(R_0 + N^2)}{\det N^2} \right]
\]

\[
I(y, s) = \frac{1}{2} \left[ \frac{\det(A(R_0 + N^2)A^T + N_\delta^2)}{\det(AN^2A^T + N_\delta^2)} \right]
\]
1D profile of optimal filters

- **high SNR**
  - reduce redundancy
  - center-surround structure

- **low SNR**
  - average
  - low-pass filter

- matches behavior of retinal ganglion cells
An observation: Contrast sensitivity of ganglion cells

Luminance level decreases one log unit each time we go to lower curve.

What is happening at low luminance levels?
Natural images have a $1/f^2$ power spectrum

- Field 1987
  - amplitude spectra for 6 images (shifted for clarity)
  - power spectra fall off as $1/f^2$
  - Fourier coefficients fall off as $1/f$

- How does this reflect the correlated structure of natural images?
- What would an uncorrelated structure look like?
- What transformation would yield a more efficient code?
Components of predicted filters

The predicted form of the optimal filter (A), is a combination of a low-pass filter (B) plus a whitening filter (C).

From Atick, 1992
Predicted contrast sensitivity functions match neural data

from Atick, 1992