Neural Mechanisms for Vision

What is Vision?

A process that produces a **description** starting from images of the external world, eliminating irrelevant information.

Humans are mostly *vision-centric*: Vision dominates our perception and memories of the world, it's the way we frame what we think.

Vision is used in two main ways:

- Object Recognition
- Guiding our movements

These separate functions are mediated by **two parallel and interacting pathways**.

We can see the brain as a system which receives *inputs* and transforms it into *outputs*.

It would be convenient to understand the nature of cognition without understanding how the brain works (Marr's hypothesis), but it's really **difficult** if not impossible **to theorise without considering neurobiological constraints**.

Vision VS Camera

- **Camera**: reproduces point-by-point representation of light intensities in one plane of the visual field.
  - More Accurate at: Quantifying absolute luminance or identifying colours.
- **Visual System**: Interprets the scene and parses it into distinct components. Separates Background from Foreground.
  - More accurate at: Classifying Objects (<200ms)

Active and Bidirectional Process

It involves more than just the information provided to the retina, it actively combines retinal information with a set of expectations stored in memory and neuronal signaling. It also relies on elements s.a. *similarity*, *Proximity*, *good continuation*.

Bayesian Theories

Visual system is seen as an *ideal observer* that uses prior knowledge to infer the **most probable** interpretation of an image.

- Prior Knowledge: Regularities governing object shapes, materials and illumination.
- Likelihood: Knowledge of how images are formed through projection on the retina.

The **more ambiguous the image** the **greater** the **prior knowledge influence** .

So perception can be either **Data Driven**, **Knowledge Driven** or both.

Visual Processing Levels

- Low Level: simple attributes (contrast, orientation, motion, depths, colour) of the visual enviroment
- intermediate Level: Low level features are used to parse the visual scene. Integration of *global contours* and *surface shapes*.
- High Level: Surfeces and contours are used to *identify objects*.

The Pathway

Retino-Geniculo-Striate Pathway:

- Retina
- Lateral Geniculate Nucleus (LGN) of the Thalamus
- Primary Visual Cortex (V1)

The Retinas

Visual Input → **Neural Signal**: *Phototransduction*.

Retinal Ganglion Cells: Output of the retina.

RGC outputs through the **optic nerve** to:

- LGN
- Pretectal Nuclei (MidBrain)
- Superior colluclus (Midbrain)

Primary Pathway: LGN → V1.

Second Pathway: Retina → Pretectal Midbrain: Pupillary Refleces that control the amount of light that enters the eye.

Third Pathway: Retina → Superior Colluculus: Controls eye movements.

Visual Field

Monocular Visions of both eyes overlap creating Binocular vision (angles depend on the species).

Usually predators have a larger monocular vision to notice their preys at a further distance, while Herbivores have larger monocular vision with an higher angle to notice threaths and run away.

Retinal Parts

- Nasal hemiretina : close to the nose.
- Temporal Hemiretina: close to the temple.
- Optic disc: Nerves that output the retinal ganglion cells axons. **Blind Spot**.
  - In *binocular vision* it gets compensated by the homologous region in the other eye.
  - In *monocular vision* the system “fills-in” the missing part f the scene using information in the surrounding areas.

Partial Decussation

Temporal Hemiretina of one eye sees the same visual hemifield of the nasal hemiretina of the other eye.

Partial decussation → Ensures that all the infromation related to each hemifield is processed by the visual cortex of the collateral hemishere.

Optic Chiasm → Fibers from each temporal hemiretina proceed to the ipsilateral hemisphere while nasal go the the contralateral optic tract.

Lesions to the Optic Chiasm cause *loss of vision of the right half of the right visual field and the left half of the left visual field*.

Phototransduction

Basic circuitry of the retina:

- **Three neuron chain**:
  - Photoreceptor
  - Bipolar cell
  - Retinal Ganglion Cell

It provides the most direct route for transmitting visual information.

Humans posses two types of **photoreceptors**:

- Rods:
  - 100 million.
  - Night (Scotopic) Vision.
  - More sensitive to light → saturated under daylight.
  - Bipolar cell receives 15~30 rods as an input.
- Cones:
  - 5 million
  - Only active for daytime (photopic) vision.
  - Less sensitive to light but have higher *Spatial & Temporal* resolution.
  - Sensitive to colour (range of wavelenghts), but mostly to green/yellow.
  - Faster response
  - Bipolar Cell receives 1 single cone.

Cones are mostly present in the **fovea**, while **Rods** are present **all over the retina**, with a sharp decline in the fovea.

Cones are then subdivided into:

- Short: Detect Blue light
- Medium: Detect Green light
- Long: Detect Red Light

In night vision green light has exactly the same effect on the visual system as red light of a greater intensity.

S cones: 10%, absent from the central fovea.

Phototransduction in Rods:

Photoreceptors contain molecules sensitive to light.
In the Dark receptors are in a state of **continuous depolarization** (*dark current*) -40mV.

When a Flashlight is presented receptors get **hyperpolarized** (-70mV).

Reason: light stimulus reduces concentration of GMPc, closing channels and making Na+ decrease.

Overall neurons in:

- Bright Regions → Hyperpolarized.
- Dark Regions → Depolarized.

RGCs response

RGCs respond to light with **action potentials** (unlike photoreceptors and bipolar cells).

RGCs constitute the **last level** of retinal processing.

They project information to the LGN through the *optic nerve*.

Different types of RGC (midget, Parasol, K) connect to different parts of LGN (Parvo, Magno and Koniocellular).

Lateral Geniculate Nucleus (LGN)

Beyond optc chiasm axons which carry nput from one hemifield join in the *optic tract* which extends to the LGN (Thalamus).

Layered structure consisting of six Layers:

- 2 **Magnocellular Layers**
- 4 **Parvocellular layers**

They are separated by **Koniocellular Layers**.

Each layers contains a map of the contralateral hemifield: the six maps are stacked on top of each other.

Layers Types:

- Parvocellular: midget RGC, ~70%, Red-Green contrast.
- Magnocellular: Parasol RGC, ~10%, Achromatic Contrast
- Koniocellular: K ganglion cells, ~10%, Blue yellow contrast.

Retina and LGN: Colour Opponency

Dimensions in the space of cone exitations produced by natural objects:

1. Luminance axis → L and M-cones signals are added

2. Red-Green opponent axis → Difference of L and M-Cones
3. Blue-Yellow opponent Axis → S-Cones signal - (L + M cones signals)

Krauspof et al., 1982 → Three independent channels in colour vision = cardinal directions of colour spaces.

Single Opponent:

- Cells increase firing rate In response of **one cone class** and increase when a different one is actvated.

Double Opponent:

- Both RFs are **Chromatically and spatially opponent**:

Primary Visual Cortex (V1)

Occipital portion of the brain.

First Level of **Information processing**.

Sensory Cortex fivided into six distinct layers.

V1 has a **visuotopic map** of the Visual Field → Upper Bank Corresponds to the lower half of the visual field.

Approximately 140 million neurons per hemisphere.

40 V1 neurons per LGN neuron.

Several Functional Categories of cells:

- Simple Cells
- Complex Cells
- End-Stopped Cells

Cortical Columns

In V1 Neurons with similar **functional properties** are found **close together** in columns (functional modules).

Specific columns for *orientation* and *ocular dominance*.

- Orientation Columns:
  - Neurons with the same orientation preference (i.e. vertical) are grouped tohether.
  - *hypercolumn* → Set of columns corresponding to a complete sequence of orientations (3/4 thousand hypercolumns)
  - *Hubel & Wiesel*: successively recorded cells shared an identical orientation of their receptive field axis.
- Ocular Dominance Columns:
  - Group neurons that respond more vigorously to stimuli presented to one of the two eyes.
  - Function: Enigma. Candidate Function → Binocular Vision, but squirrel monkeys thon have them and still have stereoacuity comparable to that of human observers.

Blobs and Interblobs

- Blobs:
  - Poorly selective for orientation. Strong preference for the **Colour** of the sitmulus.
  - Superficial Layers of V1.
  - Detectable by a specific marker for an enzyme (CO)
- Interblobs:
  - Orientation responsive

.

Cortical Columns Topology

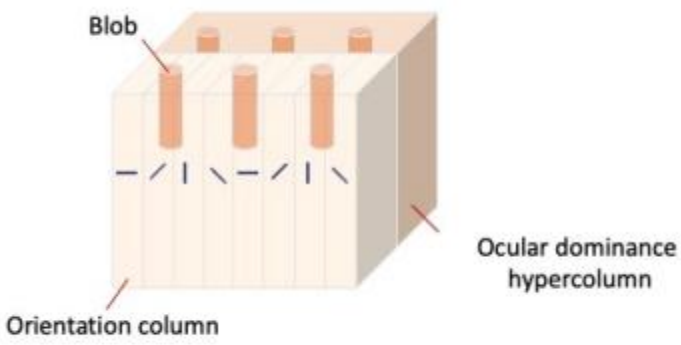
- Orientation Columns → Pinwheels
- Ocular Dominance Columns → Stripes
- Blobs → CO patches

Yacub et al: used fMRI to show the existence and spatial features of orientation selective columns in humans.
A larger Number of orientation columns are devoted to orientations around 90°.

Hubel and Wiesel: Ice Cube Model → section of V1 showing its structure.

The functional organization of V1 is based on two systems running **orthogonally** to each other:

- Orientation System
- Ocular Dominance System



Secondary Visual Cortex: V2

Thick and thin dark stripes separated by pale stripes:

- **Thick Stripes** → Neurons selective to direction of movement and binocular disparity
- **Thin Stipes** → Cells specialized for colours
- **Pale Stripes** → Orientation Selective Neurons

For every visual attribute ther must be adequate coverage of neurons with different functional properties.

There's **no generally accepted account of V2's function**. No simple properties distinguish V1 from V2.

Natural Responses to **Naturalistic Textures** Differentiate V1 from V2 in **macaques**.

Beyond V1

Set of Higher-Order visual areas organized as **Neural maps** of the visual field.

There are **two** Hierarchical Pathways, namely:

- Ventral pathway → the *What?* pathway, involved in **object recognition**
- Dorsal Pathway → the *Where?* pathway, dedicated to the use of visual information for guiding movements.

Both pathways are highly interconnected, they **share information**.
All connections are **reciprocal** → Important feature of connectivity.

Receptive Fields

Visual Neurons respond to stimuli in only a **limited region of space** → Receptive Field (RF).

How to Analyse neuron response to receptive fields?

Single cell recording (invasive, but more precise + excitation/inhibition differentiation) or fMRI. Experiments involve recording neuron activations while a a visual input (usually a light spot) is put in an animal Visual Field.

In the visual system the neuron responds to a stimulus presented in a region of space in the visual field.

Receptive fields can have different sizes based on:

- Position of the neuron along the visual pathway
- Eccentricity: RF position relative to the fovea

The amount of cortex devoted to one degree of viewing angle changes with exxcentricity (the closer to the fovea, the smaller the angle is → Higher spatial Resolution).

Retinotipy

Neuron RFs reveal an ordered organization → retinotopic (visuotopic) map.

This means that relationship between the positions of neurons in the visual areas are not randomic, but structured.

Cortical Magnification

Since regions closer to the fovea have an higher spatial resolution, bigger parts of the other areas of the Visual pathway are dedicated to them.

e.g.: The Biggest part of V1 is dedicated to 10° of the visual space.

Receptive Field properties

Properties change from relay to relay among the visual pathway.

E.g.:

- RFs of LGN are confentric, center surround receptive fields.
- RFs of V1 analyse selective contours.

RF in RGC

Concentric, Circular RFs which fall into one of two categories:

- On-Center : excited by a **light stimulus in the center** of the RF.
- Off-Center : excited by a **dark stimulus in the center** of the RF.

In general the firing rate *increases* when the *inhibitory stimulus* disappears (e.g. dark stimulus in on center).

Center and Surround reveal opposite response (*lateral inhibition*) → a uniform stimulus that activates both the center and surround causes a weak response (no variation in discharge frequency).

RGCs are **not selective for orientation of lines or edges**.

Two types of cells:

- **Transient ganglion** (M cells): spike discharge only at **start** of the stimulus.
- **Sustained neurons** (P cells): discharge **constant for several seconds**.

Lesion	Loss	Deficit
P cells	Color Perception	High spatial frequences discrimination (visual acuitiy)
M cells	Movement perception	High temporal Frequences discrimination

In general RGCs output **favours temporal changes in visual input over periods of constant light intensity**→ Image which is stabilized in the retina disappears in a few seconds.

This actually never happens because our eyes have **automatic micromovements** that refresh the input of the retina.

RGCs responde *weakly* to *uniform stimulation*, since most of the information about a scene lies in the distribution of contrast (edges).

On/Off-center resonse → Depends on *bipolar cells*.

Two classes of bipolar cells which show opposite behaviours to glutamate (Off depolarize and vice versa).

Horizontal Cells → Mediate *lateral inhibition*, they extert the influence on the photoreceptor terminals.

Lateral Inhibition enhances contrast in borders (for example when we see a discretized gradient we see the single bars changling colour although they actually don't, *Chevreul-Mach band phenomenon*).

Operating Range: RGCs responses adapt to background illumination.

Sinusoidal Gradings

Used to study visual systems.

Main attributes:

- Frequency
- Phase
- Contrast
- Orientation

Humans have a limited sensitivity to sinusoidal gradings when it comes to frequency and contrast (band pass spatial filtering).

RFs in LGN

Just like RGCs RFs.

RFs in V1

Neurons in V1 are divided into **simple** and **complex** (*Hubel and Wiesel*).

Elongated RFs and respond to a **narrow range of orientations**.

Different neurons respond to distinct orientatons (tuning curve).

Orientation → Alignment of the center-surround circular receptive fields of different LGN cells.

Simple cells → respond well to Sinusoidal Gradings of **specific frequences and phases**.

RFs in V1 are different and less homogeneous than the ones in LGN (or RGC).

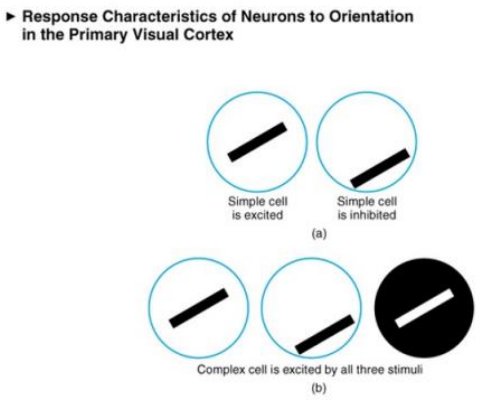
Different LGN cells converge on a Simple cell.

Complex Cells:

- Rectangular receptive fields, larger compared to Simple cells
- Respond to stimuli with *specific orientation*.
- The position in the receptive field is *not critical* → Demarcation between zones = not clear.
- React to movement particularly well.
- Selectively respond to stimuly that move in a particular direction.

Different Simple cells **converge in Complex Cells**.

Complex cells are less selective for the position of the stimulus in the receptive field → ON and OFF regions respond similarily to light over dark and vice versa.

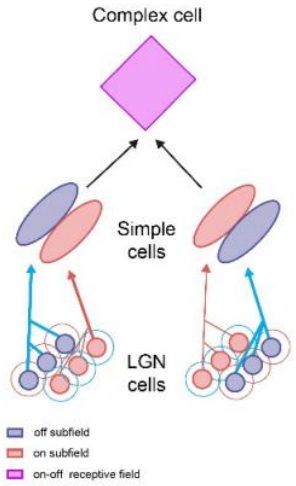


End-Stopped Cells:

- Respond better to linear stimuly of a *certain length*
- May serve to detect angles or curved lines
- ON and OFF regions of the RF have the same preferred orientation: if a straight line passes though both → no response (lateral inhibition).

Full Model

Hierarchical Model:



Hubel and Wiesel:

- **simple cell** receptive fields are constructed from **LGN inputs** with RFs aligned in the visual space.
- **Complex Cells** RFs → Convergence of Simple cells with **similar orientation preferences**.

Models (used in Neural Nets):

- *Simple cells* can be seen as linear filtering (weighted sum of image intensities with weights given by the receptive field) + Threshold.
- *Complex Cell* → Linear combination of linear filtering.

3D Vision

Based on:

- Monocular Elements
- Stereoscopic Elements

Monocular when distances > 30m:

- Familiarity
- Interposition
- Size
- Parallax
- Shadows/Lightning
- etc.

Binocular Disparity

When we **fix our eyes on a point** → convergence causes the fixed point to fall in **identical areas** (fovea).

Points outside fall in different areas of the retina → **Binocular Disparity**.

Some neurons in V1 are **selective for the horizontal disparity** of retinal images.

The same type of neurons are observed in V2, V3 and MT/MST.

MST → sensitive to stimuly that move in a **particular ditection**.

Visual Motion

- Contributes to **Object Recognition**.
- Used to establish **Depth of Distance**.
- Helps in **navigation and interaction** with the outside world.
- It serves to **direct attention**

Movement Analysis → Dorsal Visual Pathway (*Where?* Pathway)

In particular some neurons in V1 are particularly sensitive to the direction of movement.

MT/V5

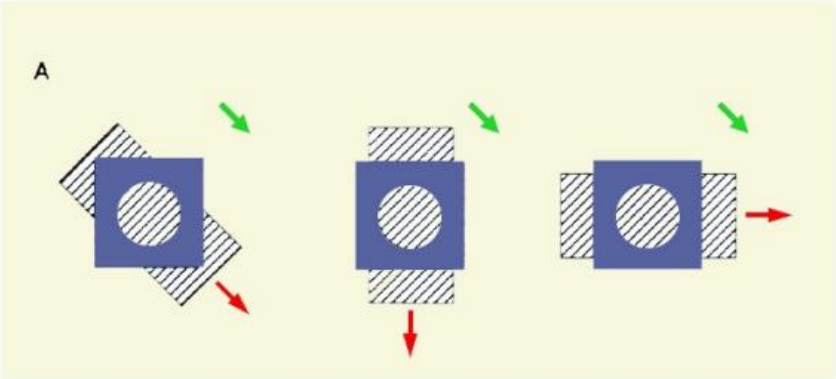
- Key role in **movement perception**.
- Small region in the dorsal path.
- 95% of its neurons respond selectively to stimuly that move in a **specific direction**.
- Similar organization to V1 (**columns**)
  - selective to direction + speed of the stimulus

Response rate:

- **Gaussian** with mean on the preferred direction (approx)

Aperture problem

Movements of a stimulus in different different directions are interpreted as movements in a single direction.



Responses are ambiguous and fail to detect the true direction of movement.

How to solve?

Two stages:

1. **Individual components** are analyzed → loval components perpendicular to the movement of the contour
2. Higher order neurons **integrate** local components.

Integrated signal = Real Direction.

Eample: Moving Grid Pattern has 2 diagonal Components.

MT integrates the two different components and responds to a singular direction.

MT is also able to responde to movements in opposite directions.

Selective Visual Attention

Can be:

- Directed toward a **Specific region of space**
- Guided by **non spatial properties** (i.e. colour, shape, direction of motion, etc.).

MT is crucial for the perception of Visual Motion.

Spatial Attention modulates the response of neurons **selective to direction** (MT/MST).

Experiment 1: **Effect of spatial Attention**

“When a monkey was attending a moving stimulus inside a receptive field the neurons have an increasing response”

Experiment 2: **Effect of non spatial, featural attention**

Authors compared neural responses when the stimulus was moving in a direction or another. → Attending to the preferred motion outside the receptive field increased the response by 13% on average.

Attending to a given direction **enhances** the responses of neurons whose preferred direction **aligns with the attended one** → Spatial and Feature based attention represent **Separate and summable processes**, they have **multiplicative effect on neuron responses**.

V4

- Intermediate Cortical area in **Ventral Visual Pathway** (*What?* pathway).
- Crucial for **Visual Object Recognition**.
- Primary Role: facilitate **segmentation** of the visual scene.

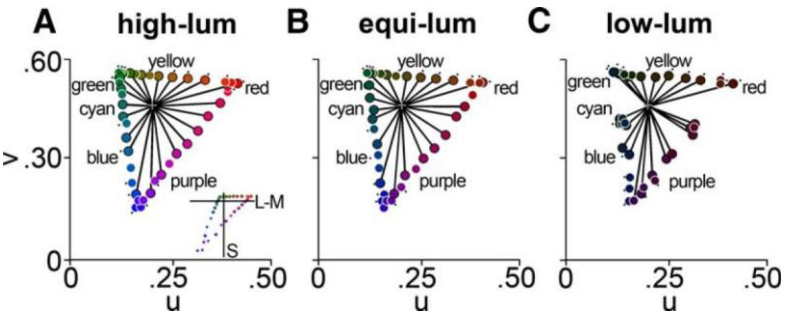
V4 is **not an homogenous area** → Many functional domains = **Globs**

- **Globs**: Colour Processing
- **InterGlobs**: Orientation and Shape Processing

Response to Colour

Colour Voxels are better activated my **equiluminant** colour (No white/black grating).

Equiluminant Colours can be mapped into a **polar map** in which angles represent shades of colours.



Hue responses can be compressed into Polar Plots.

Luminance invariance does not imply that luminance does not modulate the responses → it just does not drastically shift or cancel hue tuning.

Interglobs → Do not show strong hue tuning.

Colour Constancy

Effect:

- The **perceived or apparent colour** of a surface **remains constant** despite **changes in the illumination** (intensity/spectral composition).
- Photographs need **White Balance** to perceive those changes

In general:

Brain determines an objects's colour from a **local comparison** of the light reflected from the object and the **adjacent areas** of the scene.

Retinal luminance problem

Retinal Luminance = product of *Illumination* and *Reflectance*.

These physical determinants of retinal luminance values are conflated in visual stimuli and **can't be disentangled by any algorithmic process**.

(*Note: Same as CV reflectance problem*)

Object Identification and Categorization

Visual experience → **centered on objects**.

Visual Object = Set of Charateristics (features) grouped or joined in descrete units on some basis (proximity, similarity, closure etc.).

Visual Recognition = Assign a *verbal label* to objects in the Visual Scene.

Two Main types of Objects Recognition Tasks:

- Identification: Easier for computers, e.g. "I's a Siamese Cat"
- Categorization: Easier for Humans, e.g. "I's a Cat"

Categorization: Two objects or events are treated equivantely. E.G. when two objects are labeled the same way or the same action is applied to them.

“Although the stimuli are distinct, organisms do not treat them uniquely; but they respond on the basis of past experience and categorization.”

We **efortlessly and rapidly** (100-200ms) detect and classify objects among tens of thousands od possibilities → Basic functions of living beings.

Object Recognition = **Integration of visual feaures extracted at earlier stages in the visual pathways** → Generalization across different members of an *Object Category*.

Selectivity and object Invariance

Computational difficulty of Object Recognition is requiring both:

- Selectivity: different responses to distinc objects, s.a. two different faces
- Invariance with respect to image transformations, s.a. rotations of a face.

Perceptual constancy (Object invariance problem) → **Really easy** to deal with **for humans**, but main **obstacle** in *Computer Vision*.

Brain areas dedicated to O-R

Inferior temporal cortex → large region of the brain that includes at least **two major functional subdivisions**:

- Temporal-Occipital (TEO) area
- InferoTemporal(IT) area.

Both of them are located in the **temporal lobe**.

PIT CIT AIT:

In **contrast** to V1, V2, V4 → **No Clear Retinotopy**.

Functional Columns in IT

Tanaka 1996 → suggested that responses to effective stimuli are organized in topographical order:

- **two adjacent neurons are more likely to respond similarily to a set of stimuly than distant ones**.

  The neurons that respond to different parts of an object are **not randomly arranged in IT**.

  Response to the same Elements → Same Cortical Column.

  Tanaka Also Hypothesized that the Representation of each Object occurs **Not through the activation of a single neuron**, but **by a population of IT neurons** which encode features of the object distributely (Object components).

  The Majority of IT neurons respond to a stimulus **only when it's presented from a specific points of view**.

  10% of them responde to stimuly regardless of their position with respect to the observer (**View Independent responses**) → This means that IT is able to form a **relatively abstract** representation of an object.

  Gnostic unit → Neuron that can recognize a *Complex Object* (Hierarchical Model from Features to objects). Idea: Signals the presence of a complex and Highly specific stimulus.

Local or distributed coding?

Grandmother cell: Researcher's hypothesis of a **Gnostic unit** that only activates when we see a grandmother.

- **Distributed code Hypothesis**: Object recognition is the result of a distributed activation pattern on the population of IT neurons.
  - It could easily explain why we can recognize similarity between objects (e.g. tiger and a lion) and make mistakes between visually similar objects.

  The results of the studies on single neurons of the temporal lobe are in agreement with the theories of the distributed code of object recognition.

  Although some cells can recognize complex Objects (relatively, not absolutely).

IT neurons response

- IT neurons:

  - respond **only to visual stimuli**
  - their RF tend to be **LARGE** → can generalize

- Encode **Complex Charateristics** of the stimulus (e.g. crescent of a particular colour and texture).
- Complex IT neurons provide **INPUTS to higher order neurons** that respond to specific objects.

- IT neurons response examples:

- Faces → Distributed Pattern (No Grandmother's cell)
- Emotional Expressions
- Direction of Gaze
- Hands

- IT cell selectivity is usually invaiant to:

- Size
- contrast
- colour
- exact location on the retina

- The **activity** of IT neurons can be **Animal's attention** (IMPORTANT)

- IT cells exhibit both **short** and **long term memory**.

Response Modeling

The **response** of a population of Neurons to a **view of an object** can be considered as a response vector in an **N** dimensional space (N = number of neurons in the population).

Experiment: small Neuronal population response has been classified over short time intervals: it has shown robust information about "identities" and "cateegories" of objects.
Readout: Training a regularization classifiar to learn the neuronal responses map to each object label and predictiong them.
This classifier's performance increased LINEARLY with the number of sites → indicative of a distributed representation (not grandmother like). (Hung et al 2005).

Object manifolds are thought to be gradually untangled though **nonlinear selectivity** and invariance computations applied at each stage og the **ventral pathway**.

Higher rates of visual processing → neurons maintain selectivity for objects across changes in view: more flat and separated (untangled).