Wavelets in vision

Foundations of vision

Topics

- Part 1
- Vision sciences background: Foundations of vision
- Image Representation and sparse bases
- Multiresolution representations

- Part 2
- Color vision and colorimetry

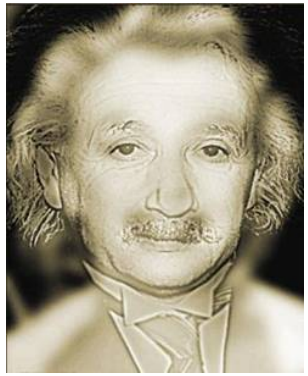
Foundations of Vision

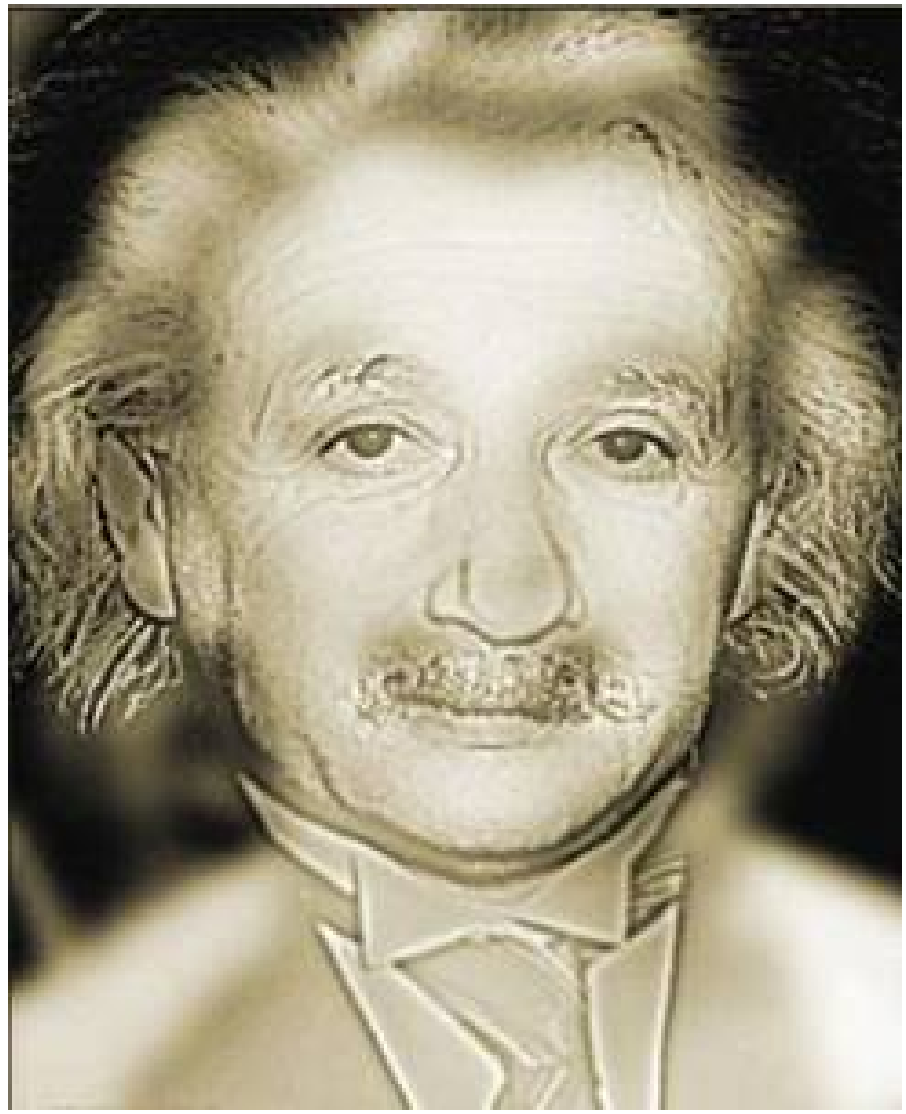
Part 1

Multiresolution in vision

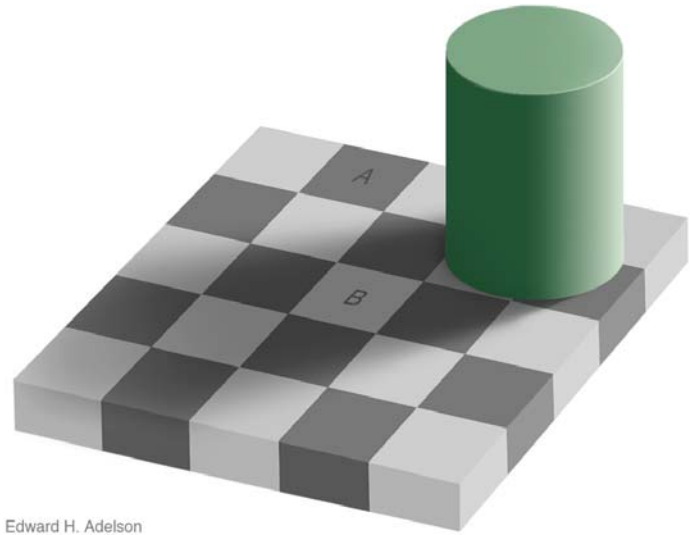
Can you believe your eyes?

Can you believe your eyes?

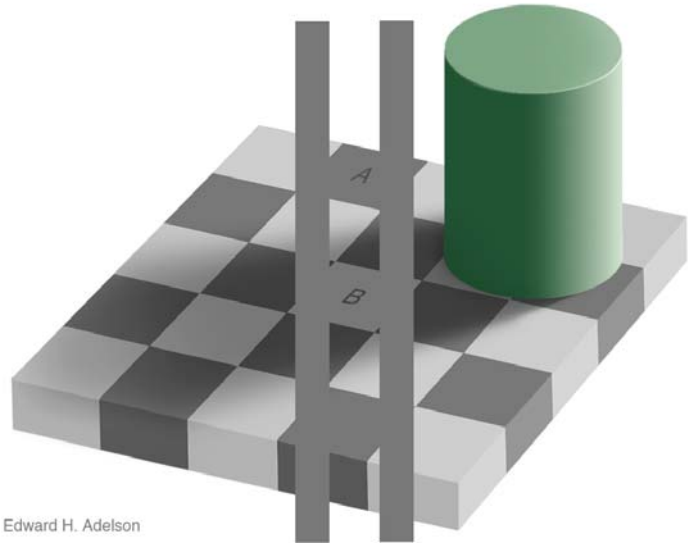




Can you believe your eyes?

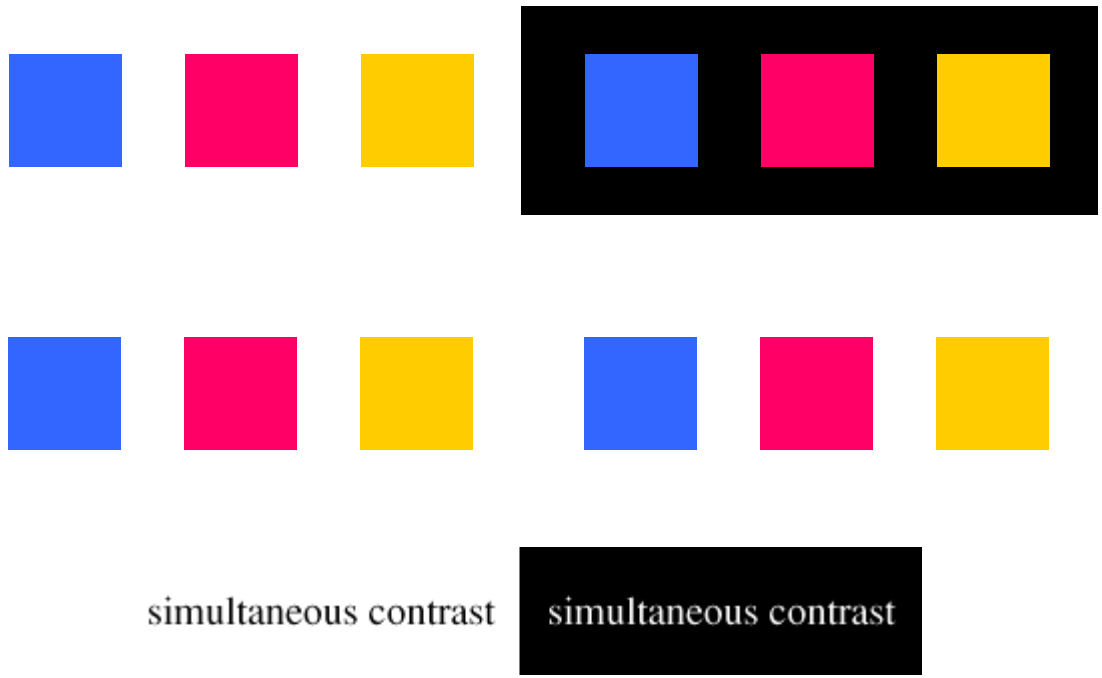


Edward H. Adelson



Edward H. Adelson

Simultaneous Contrast



Simultaneous contrast describes the way a color object seems to change size and intensity based on the colors nearby.

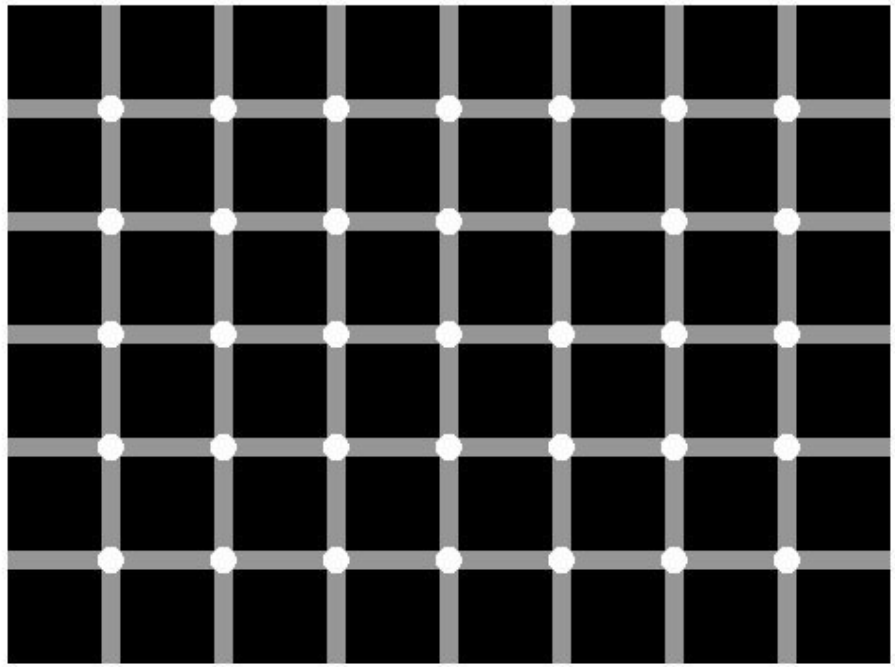
Chromatic Induction

Chromatic induction describes the way adjacent colors alter the way the color itself is perceived

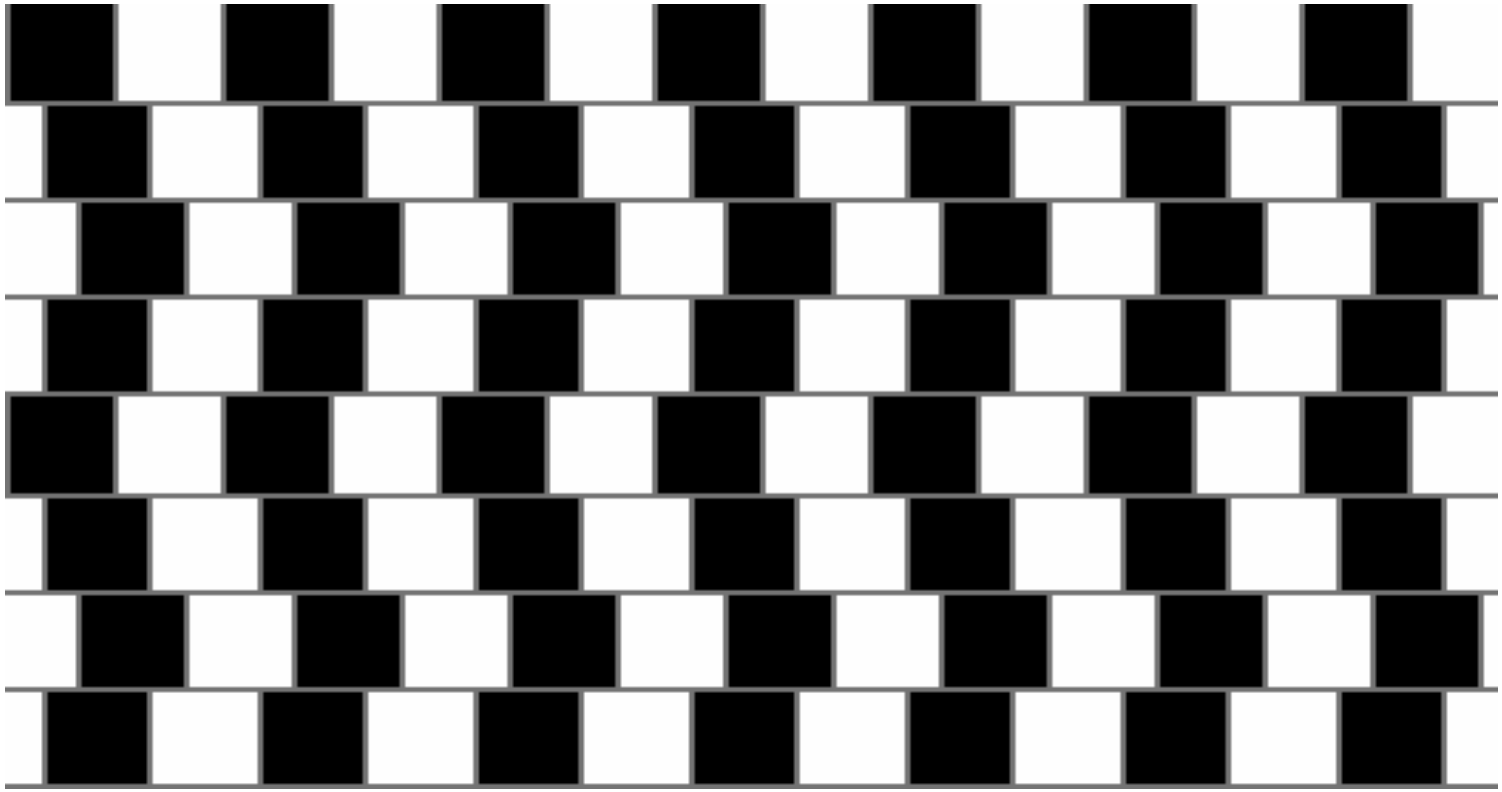
what color is this?

Afterimage

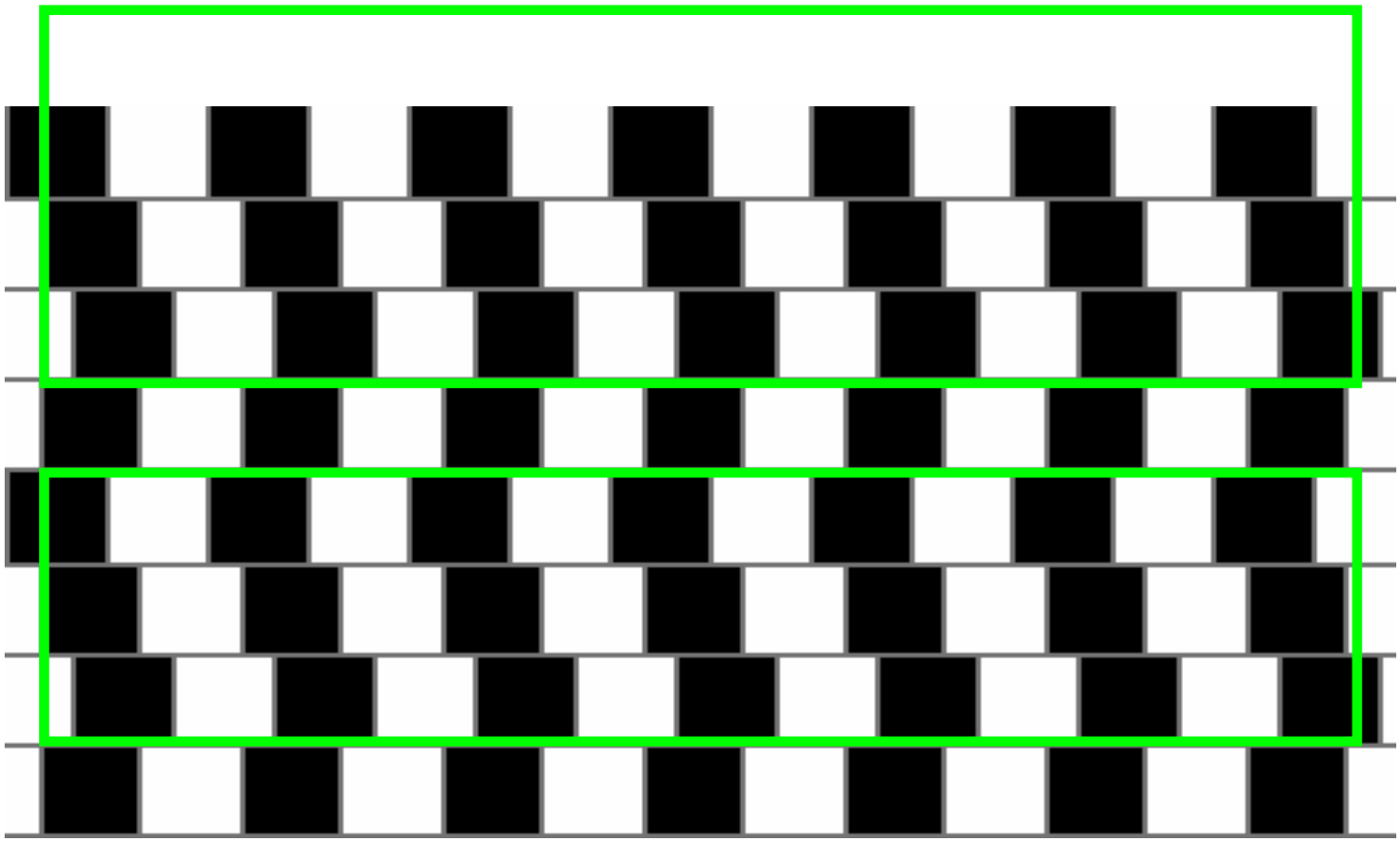
Can you believe your eyes?

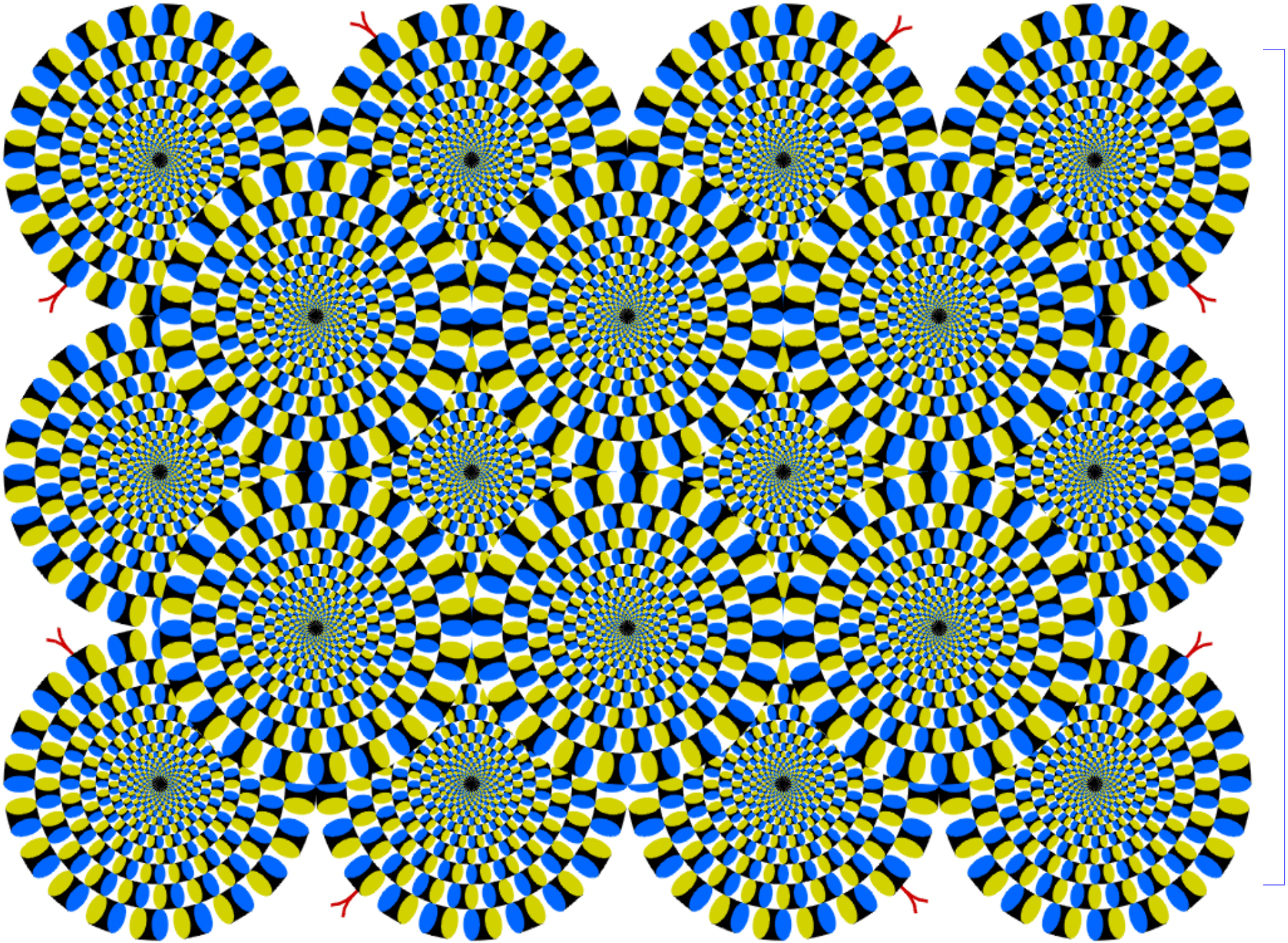


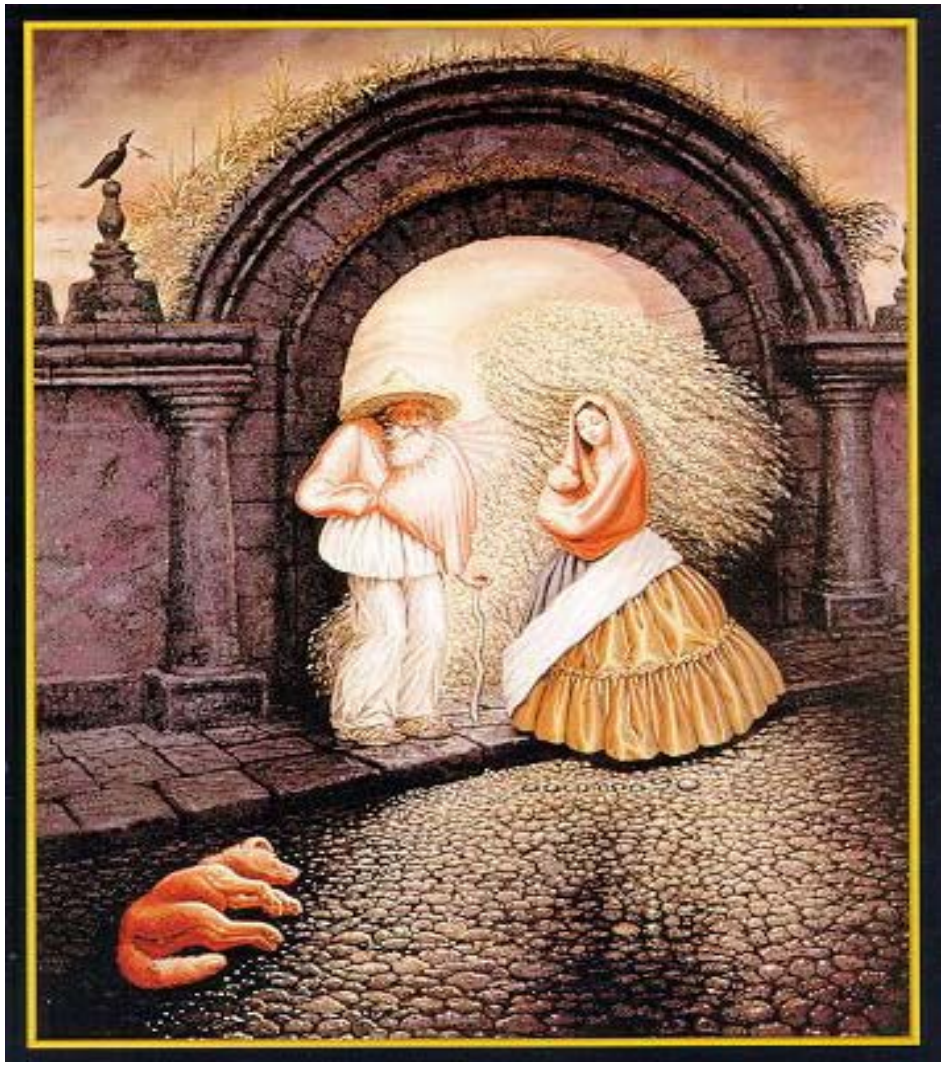
Can you believe your eyes?



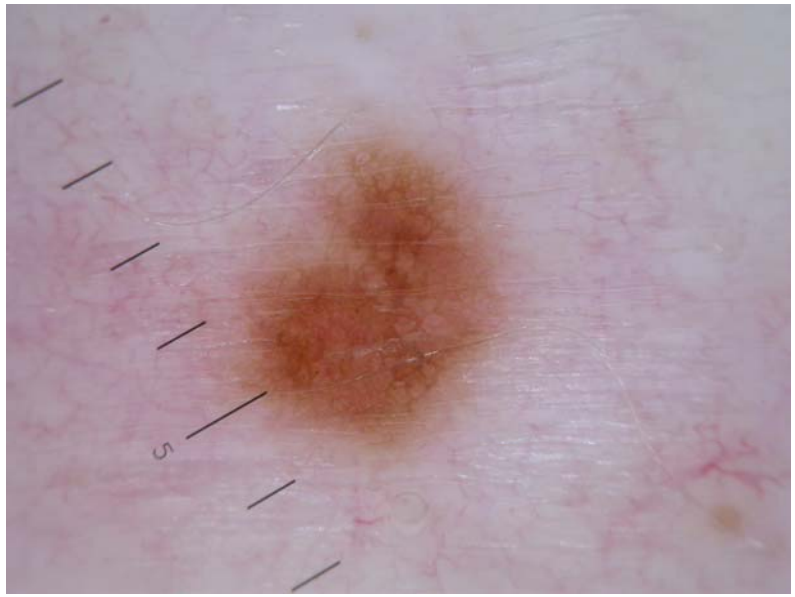
Can you believe your eyes?



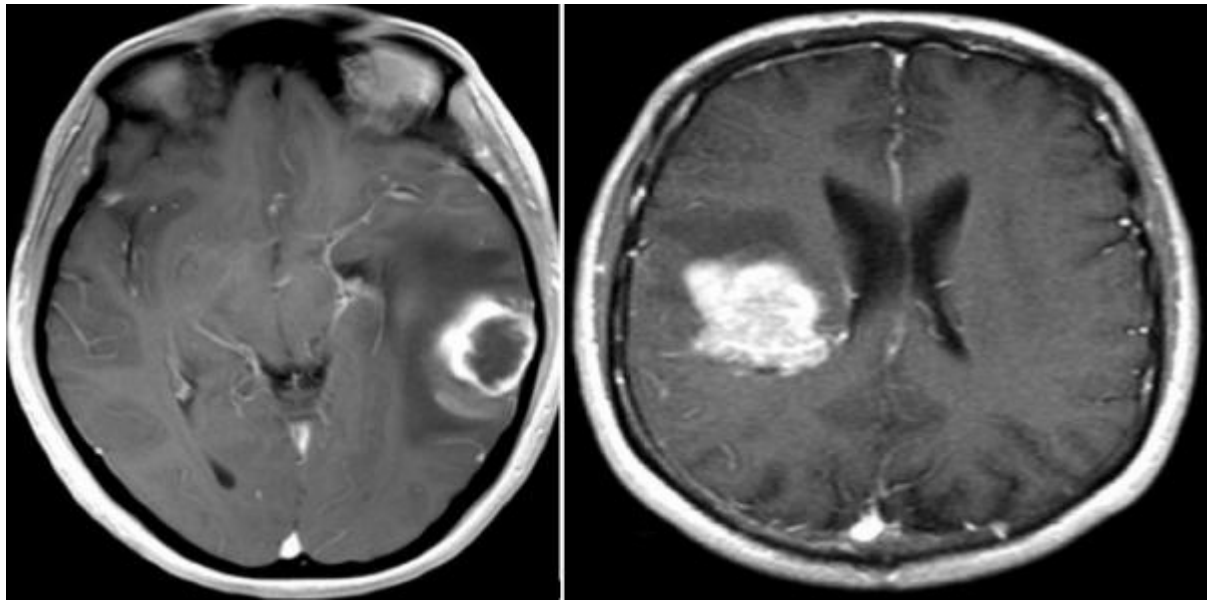




Relevance



Relevance



Take home message

- There is no direct mapping between what you *perceive* when looking at an image and what is *encoded* in the image. The link between the two can be made explicit by a model
- Multiresoultion plays a role in perception
- Color vision is not straightforward neither is color imaging
- Image interpretation is ambiguous and depends on prior knowledge and expectations

Understanding vision

A problem of reverse engineering!

How to study Vision (by Wandell)

- Encoding
  - How the retinal image is encoded in the visual pathway
    - Determines the quality of the information available to higher levels of processing
    - Sets the benchmark for performances of automatic reproduction systems
    - Nice illustration of the complementarities between physical calculations, biological experiments and behavioral studies

- Representation
  - How the encoded image is represented by the neural response within the peripheral and early cortical visual pathways

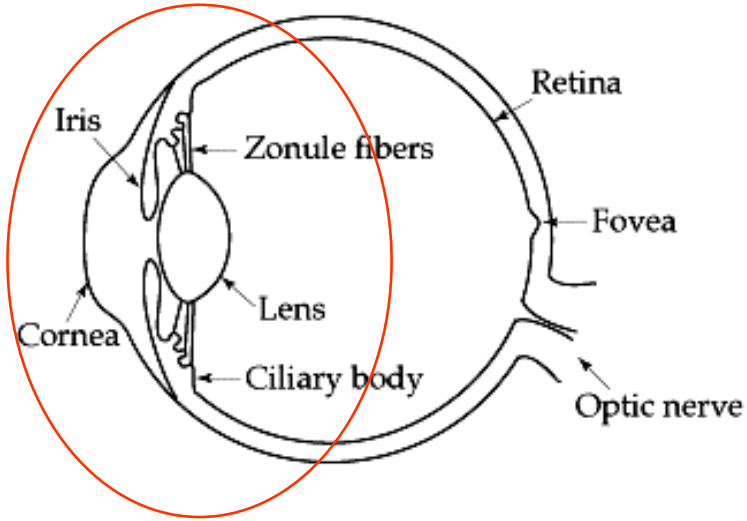
- Interpretation
  - Perception is an *interpretation* of the retinal image, not a *description*
    - Assigns perceptual properties such as color, motion and depth

How to study Vision (by Wandell)

- Since the information provided by the retinal image is imprecise, image interpretation is an *inference process*, which is possible due to the *statistical regularities* of the natural environment (and thus of the retinal image).
- Understanding such regularities and how to use them to interpret the retinal image are central to vision science.

Part 1: Encoding

- Low-level processing
- Image formation
  - Optical quality of the eye

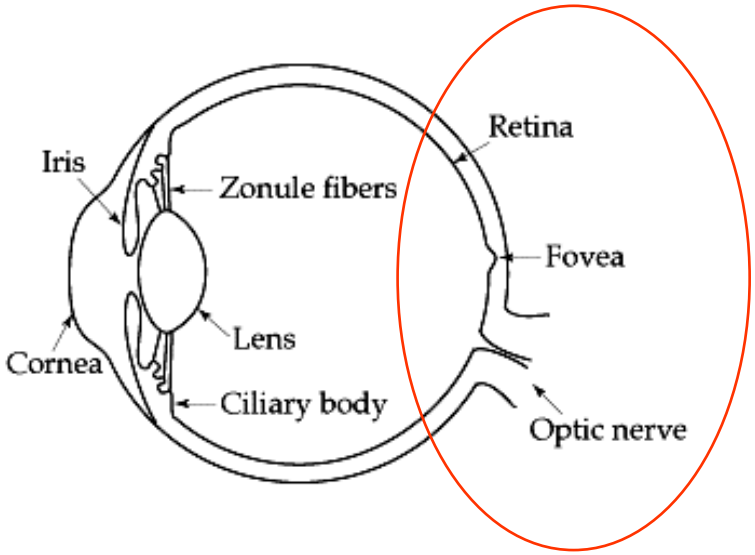


- Photoreceptor mosaic
  - Photoreceptor types
  - *Geometrical* aspects
    - The cone mosaics
    - Sampling and aliasing

- Wavelength encoding
  - Scotopic and photopic conditions
  - Basic issues on color vision: Color matching functions

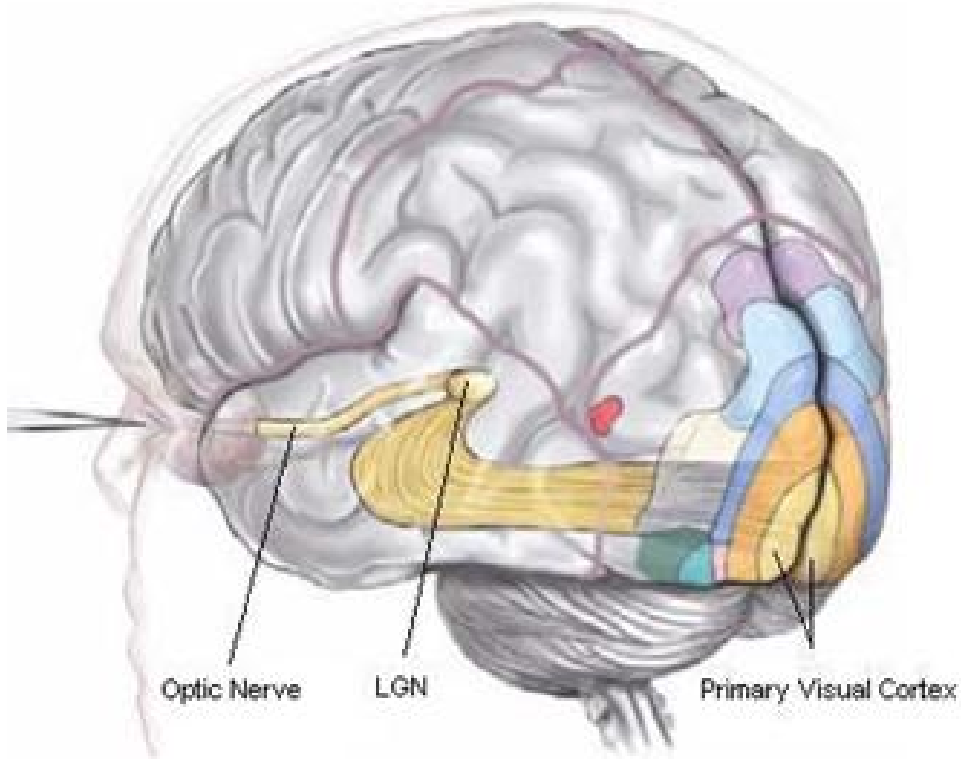
Part 2: Representation

- Low-to-mid level processing
- The *Retinal* representation
  - The retina
  - Retinal Ganglion Cells response to light
    - Receptive field
  - Light adaptation
- The *cortical* representation
  - The Visual Cortex
  - Receptive fields in the visual cortex
- Pattern sensitivity
  - (Spatial and Spatiotemporal) Contrast Sensitivity
  - Pattern adaptation
  - Masking and Facilitation



Part 3: Interpretation

- High level processing
  - Color *vision*
  - Motion and depth perception
  - Seeing

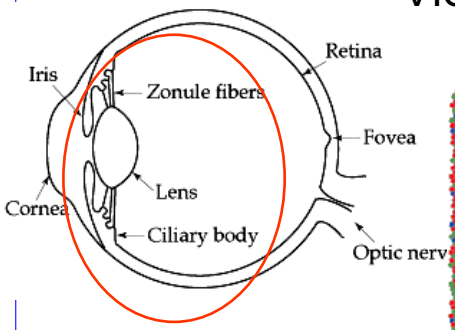


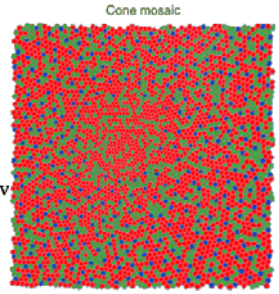
Encoding

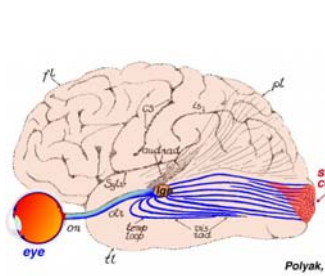
- Process that describes how the light (photons) entering the eye are captured by the photo-detectors present in the retina and the resulting retinal representation of the visual stimulus (image)
- Issues
  - Limits set by the optic of the system
    - Cornea, lens
  - Sampling
    - The photoreceptors form a discrete grid, rising the classical sampling related issues
  - Normalization
    - The dynamic range of the output of the visual neurons is finite and much smaller than that corresponding to the change in illumination experienced in a typical day (light *adaptation*)
  - Transfer function
    - The sensitivity of the photo-sensors depends on the frequency of the stimulus and on the illumination condition
- Assumption: the image formation process is linear
  - Can be investigated and modeled by the linear system theory

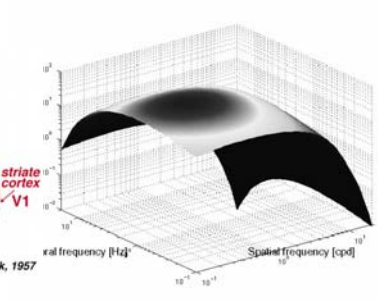
Visual and Imaging chain

VISUAL CHAIN

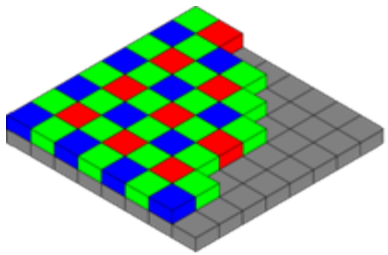








Optics	Sampling (A/D)	Multiscale representation	Filtering	Processing



IMAGING CHAIN

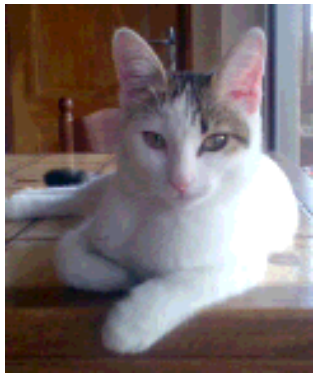
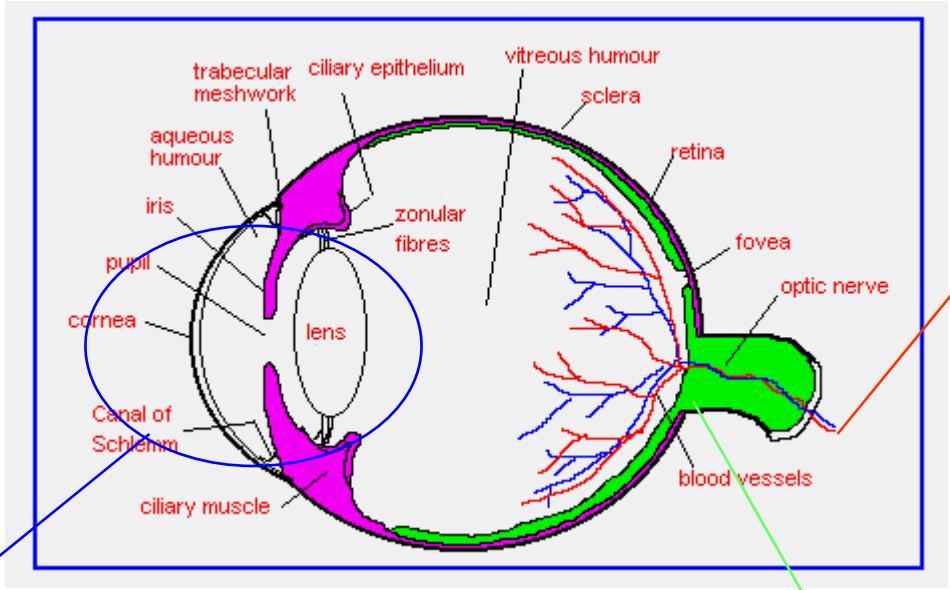


Image formation

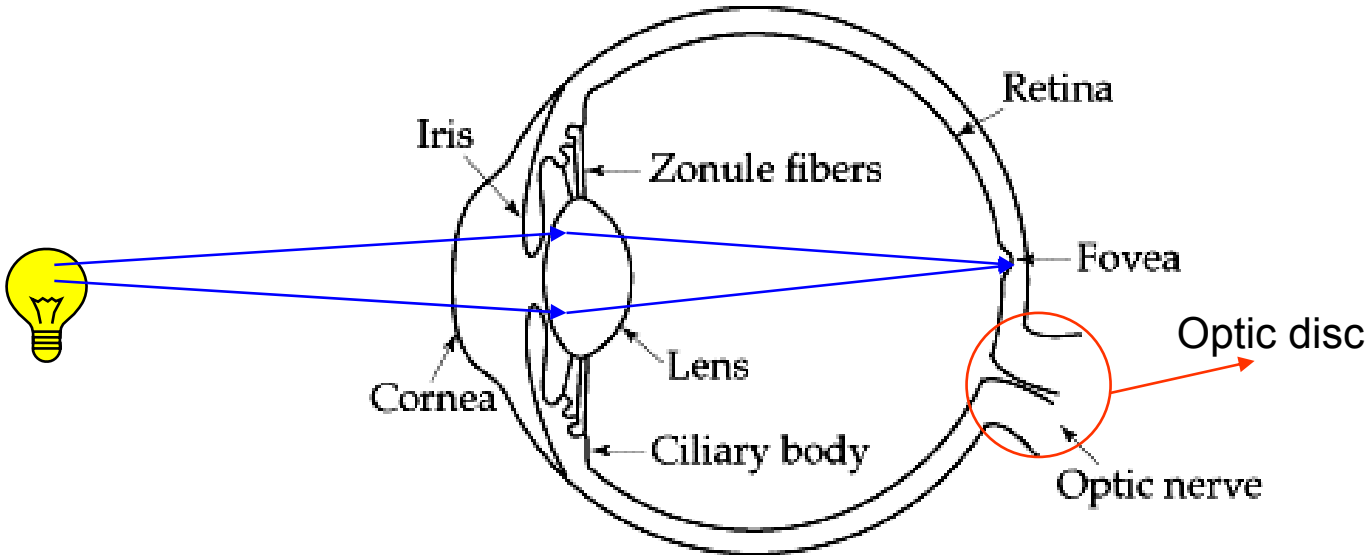


The neural responses are transformed into neural representations within the optic nerve which brings them to the brain to form other **cortical representations**

Cornea and lens focus the impinging light to the retina

The **photoreceptors** on the retina transpose the quanta into neural responses

Image formation



- The light entering the eye is brought to focus on the retina by the eye optics (cornea+lens)
  - The focus of the optical system must be kept on the retina in any condition
- The retina is a thin layer of neural tissue. It consists of different types of neurons. The axons of some of these are collected into the optic nerve.
- The optic nerve exits the retina at the *optic disc* to bring the signal to the brain for further processing

Experience your black spot

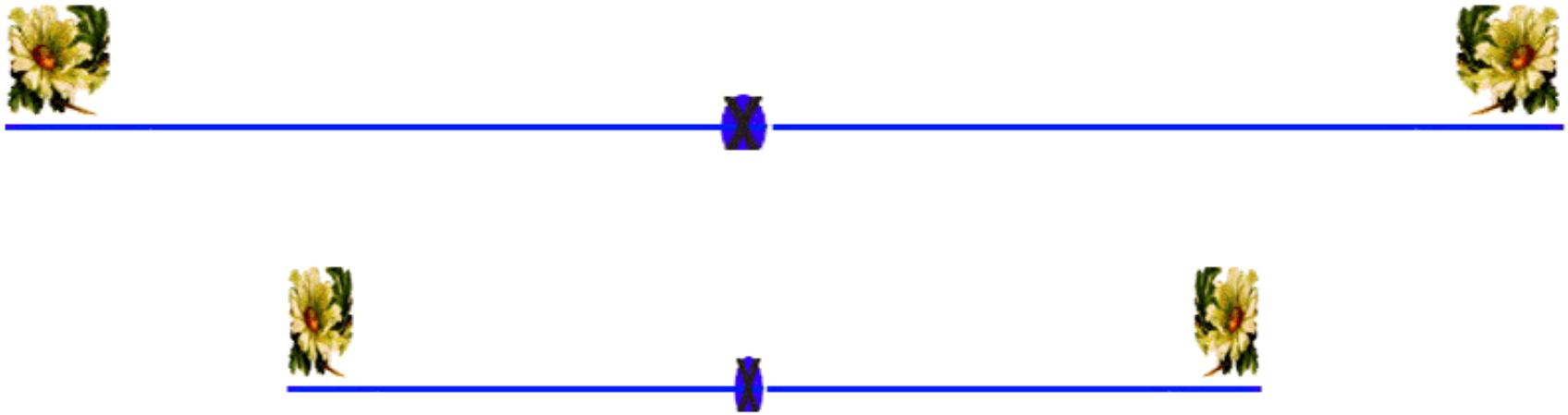
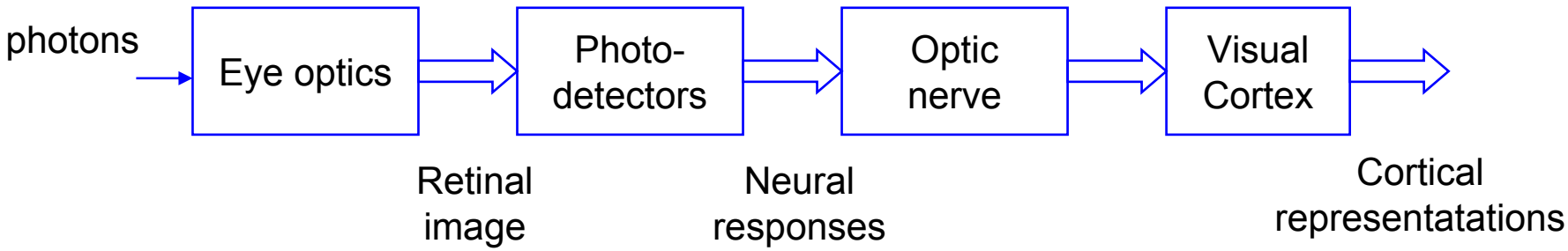


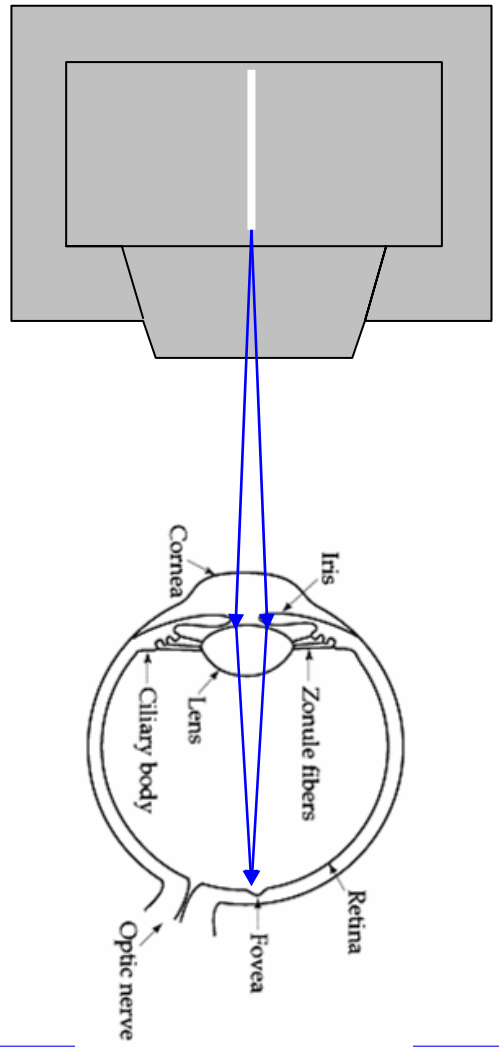
Image formation

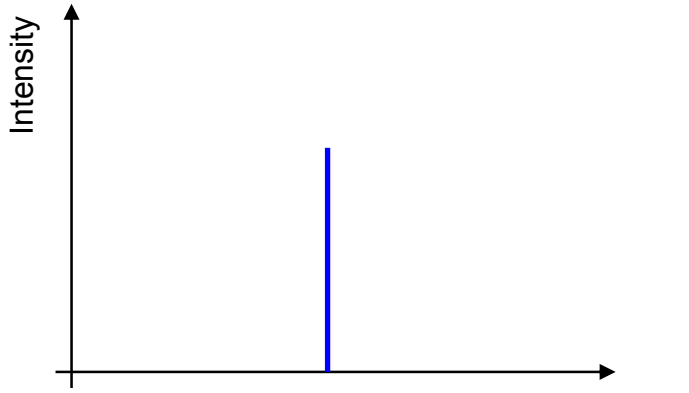


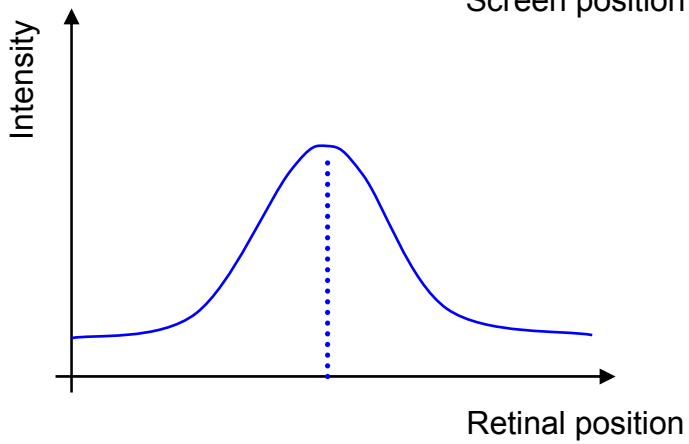
Due to the complexity of the optical system, it is reasonable to expect that the response will be characterized by a *line spread function*

- The retinal image is a blurred version of the original

Retinal image formation

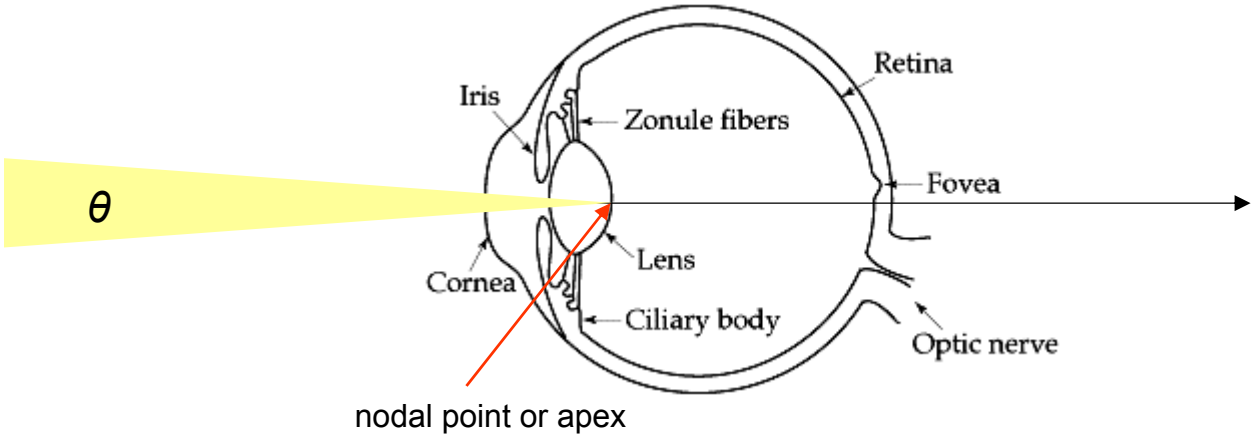






The retinal position is measured in eccentricity or units of *visual angle*

Retinal position

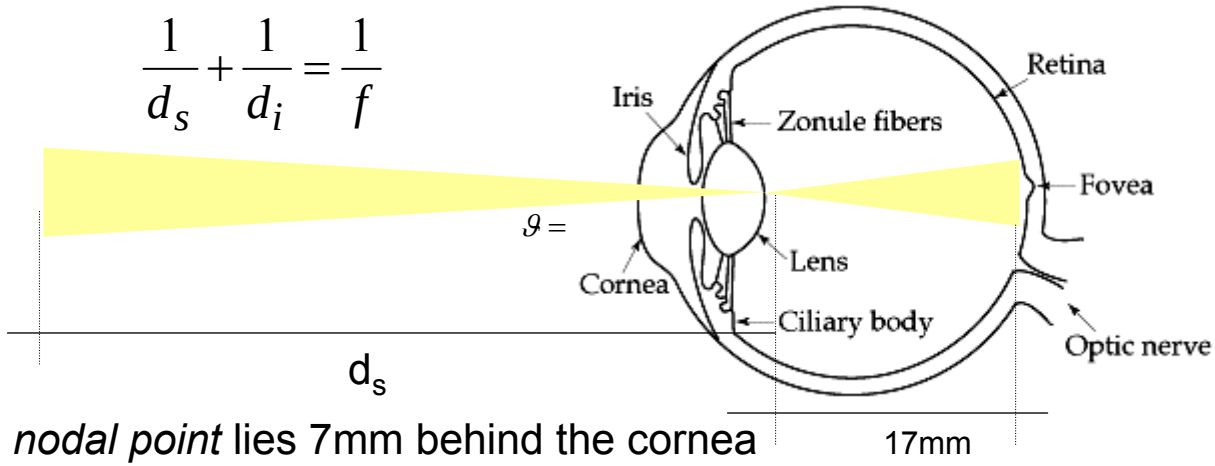


- *Eccentricity* is the measure of the distance on the retinal surface, which is either express as the distance in mm on the retinal surface or (more often) as the *visual angle* between some point on the retina and the *center of the fovea*
- Objects with the same visual angle have the same size on the retina

The optics of the eye

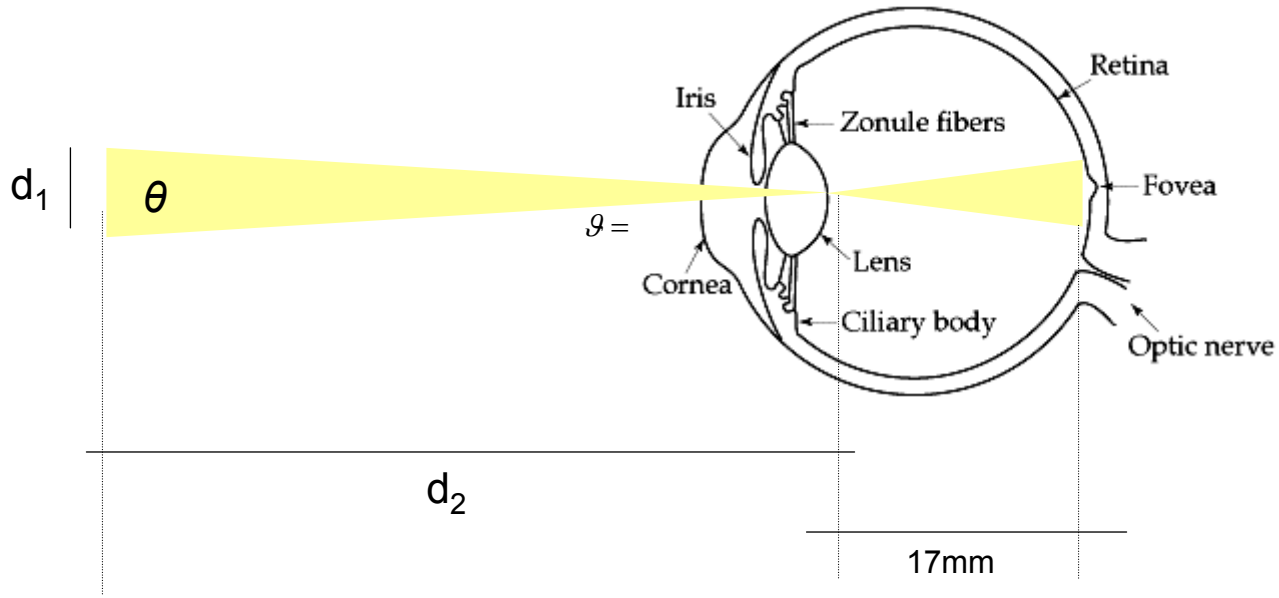
- Geometric optics approximation

$$\frac{1}{d_s}+\frac{1}{d_i}=\frac{1}{f}$$



  - For most eyes, the *nodal point* lies 7mm behind the cornea
  - The distance from the nodal point to the retina is the ***posterior nodal distance***, and it is about 17mm for adult humans. This must corresponds to the ***focal distance*** of the system.
  - To have the image at focus on the retina, *di*=17mm. The *optical power* of the system (*p=1/f*) is thus dynamically changed to match such a condition when changing the distance of the object (screen) from the eye. In practice, the image of the object is focused on the retina for *ds*>=1m.
    - Such optic power corresponds to 58.8 dyopters

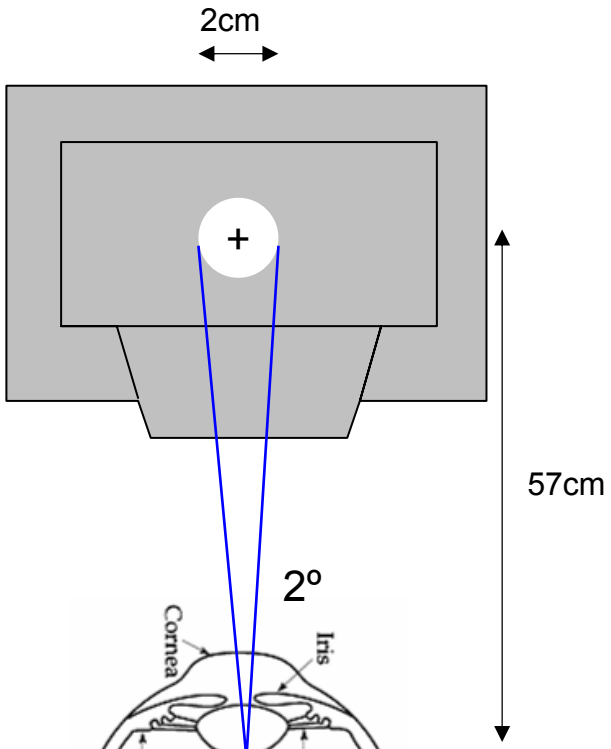
Eccentricity

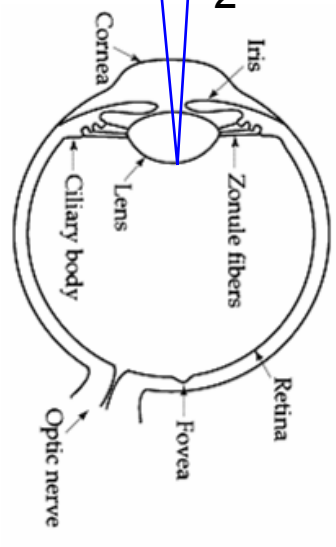


$$\vartheta = 2\arctan\left(\frac{d_1/2}{d_2}\right)$$

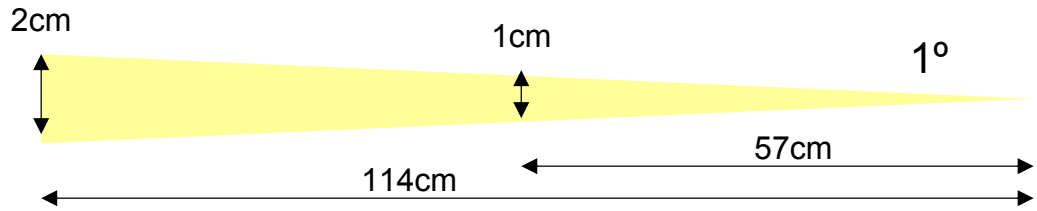
- This allows to set the size of the stimulus on the screen such that the corresponding retinal image covers a pre-defined region of the retina, for each distance between the screen and the subject

Eccentricity





- Some typical values
  - The **fovea subtends 2 degrees of visual angle**
  - To stimulate the fovea the stimulus must be centered on the screen and cover a visual angle of 2 degrees
    - 2cm on the screen at 57cm
    - 4cm on the screen at 114cm
  - By going farthest at fixed stimulus size the visual angle decreases and different portions of the retina are stimulated

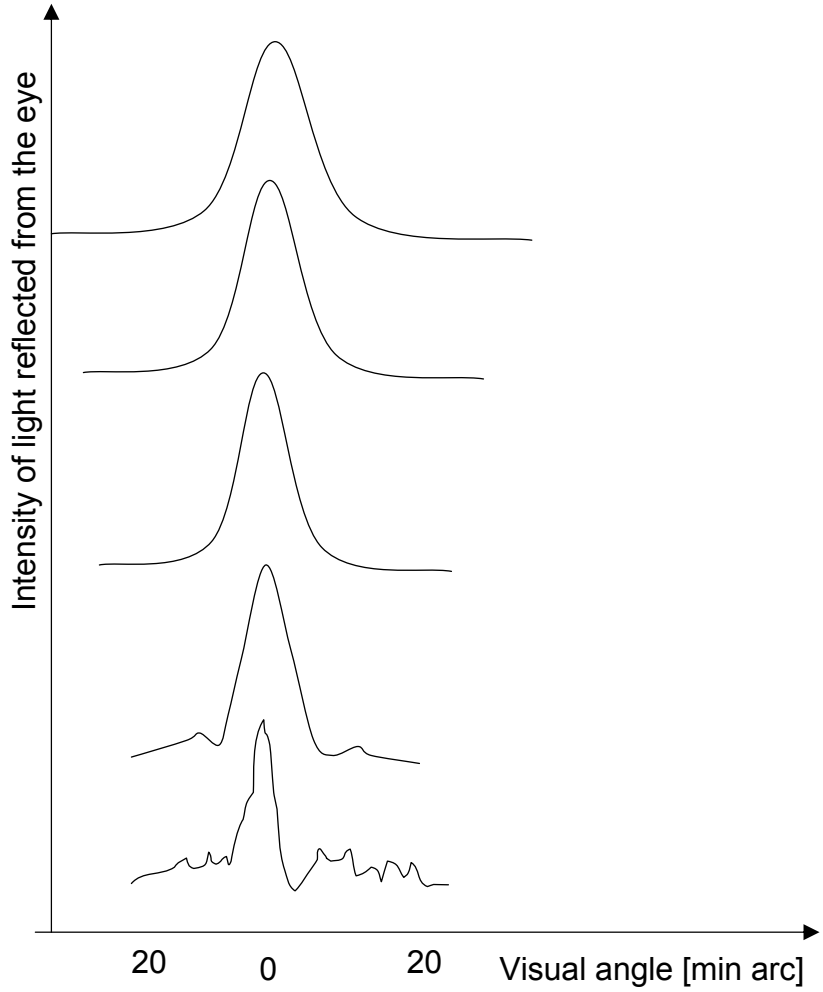


$$\arctan\left(\frac{1}{57}\right) = 1^{\circ}$$

Linespread function

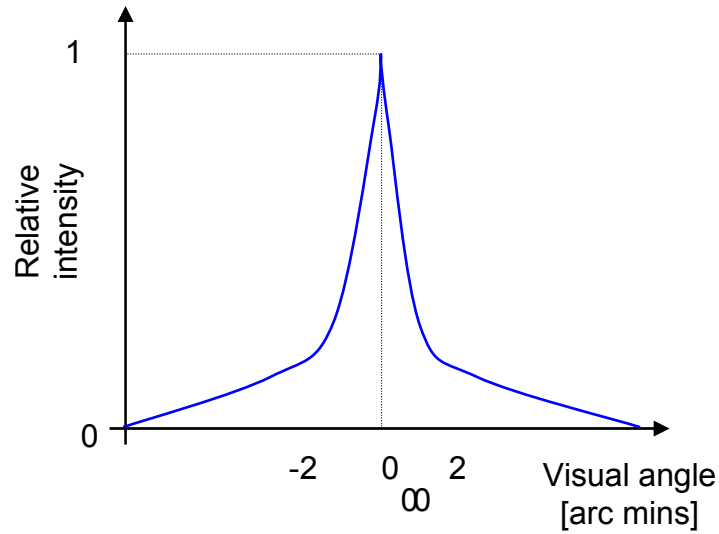
- Assumptions [Campbell and Gubish]
  - Linearity
    - Homogeneity and superposition hold
  - Shift (translation) covariance
    - A shifted input generates a shifted output
  - No phase shift
    - The phase of the retinal image is the same as the phase of the stimulus
  - The attenuation of the signal is the same in the two directions
  - Stimulus composed of a single vertical line on the screen at different positions
  - Different illumination conditions (pupil aperture)
- The estimated linespread function
  - Is bell-shaped
  - The width (blurring) depends on the illumination
    - When the pupil is wide open the width of the lens increases and the amount of blurring (defocusing) increases

Linespread function

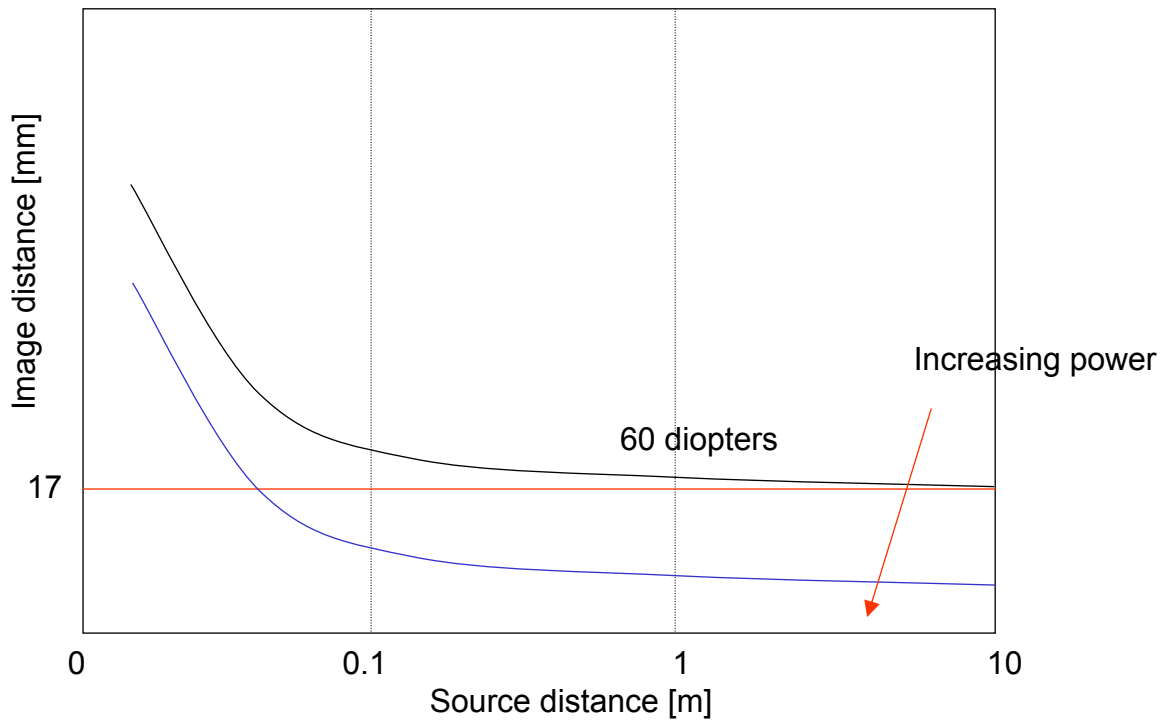


- Westheimer's model (1986)
  - Under certain viewing conditions and for a pupil diameter of 3mm

$$l_i = 0.47e^{-3.3i^2} + 0.53e^{-0.93|i|}$$

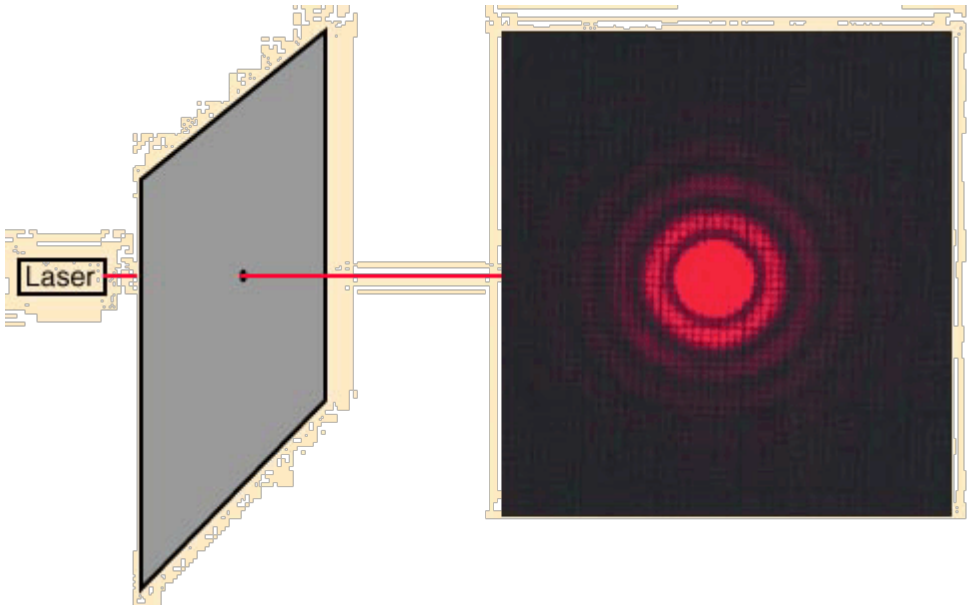


Depth of field



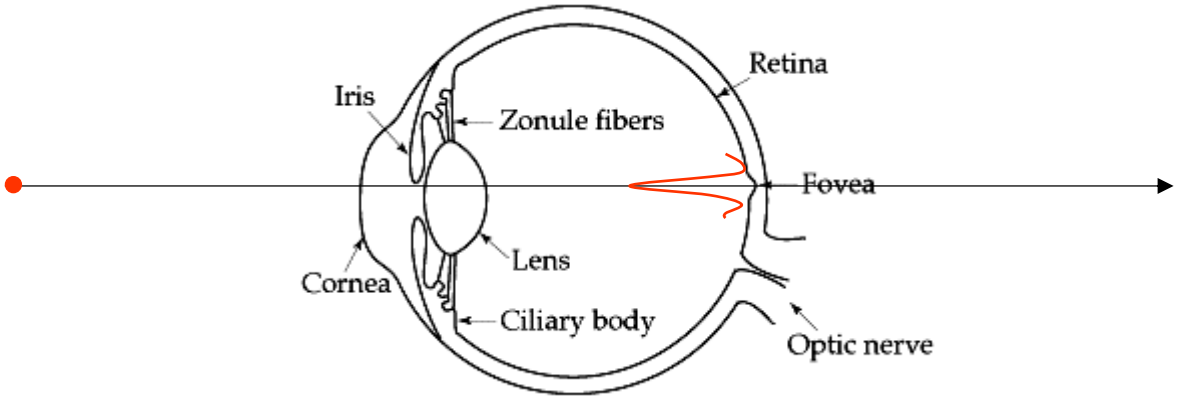
Diffraction

- Diffraction takes place because the light passes through the circular aperture defined by the pupil.
- For apertures as small as 2mm, the quality of the optical system is quite high (small portions of the cornea and lens near the center of the visual field). In this conditions the main source of image blur is diffraction.

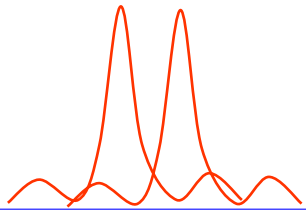


Pointspread function (PSF)

- Pointspread function (PSF): Image of a point source of light on the retina
  - Aberration
  - Diffraction

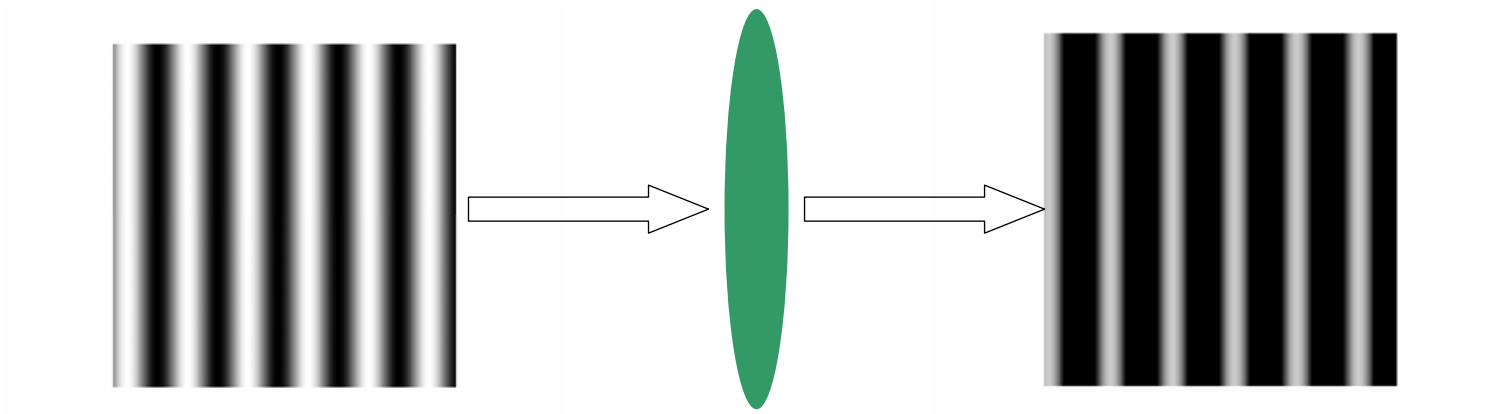


  - Spreads the response to one single spot to about 20 cones
  - Limits visual acuity
    - Points can be resolved up to the point the peak of the one falls into the trough of the other



Other characteristic functions

- Optical Transfer Function (OTF): Complex function that measures the lost in contrast and the phase shift of a sinusoidal target
  - Modulation Transfer Function (MTF)
    MTF=|OTF|
  - Phase Transfer Function (PTF)
    PTF=phase{OTF}



Photoreceptor mosaic

- The retinal image is sampled by the photo-receptors of the retina
  - Discrete sampling grid

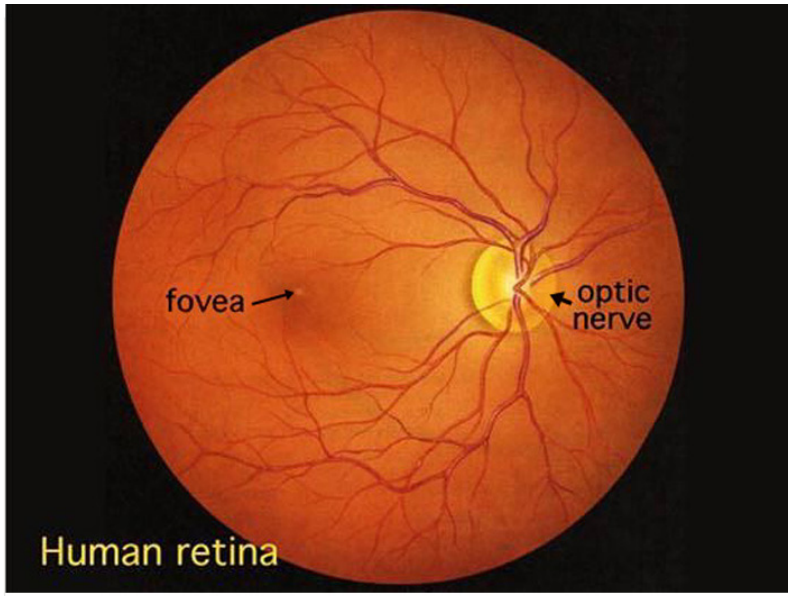


Fig. 1. Human retina as seen through an opthalmoscope.

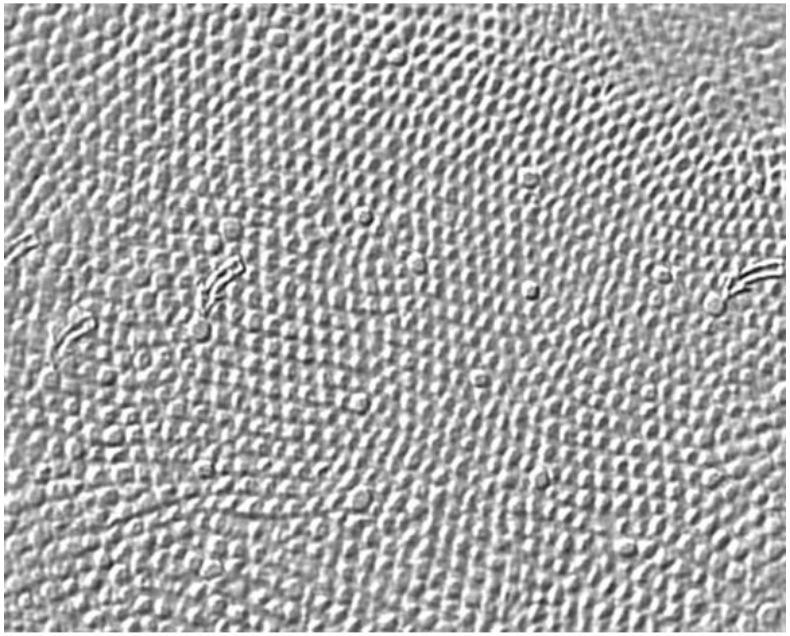
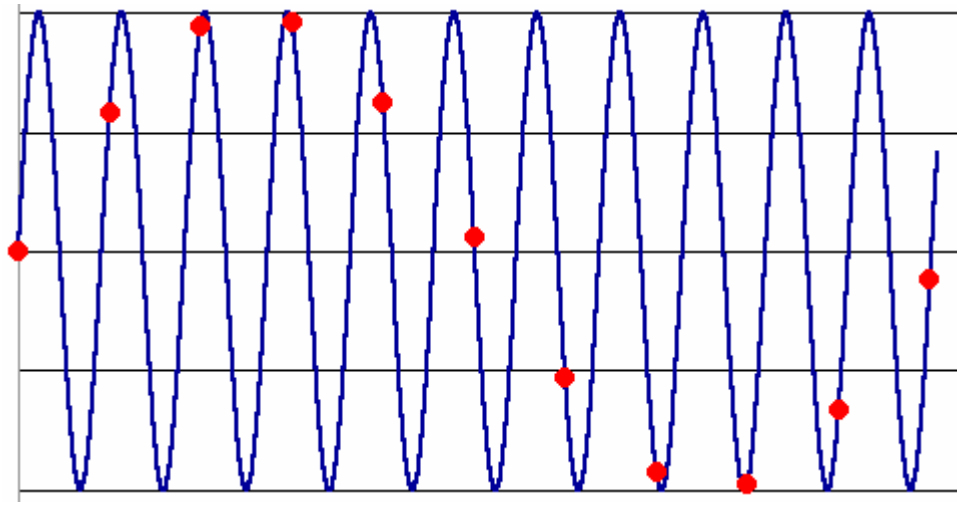
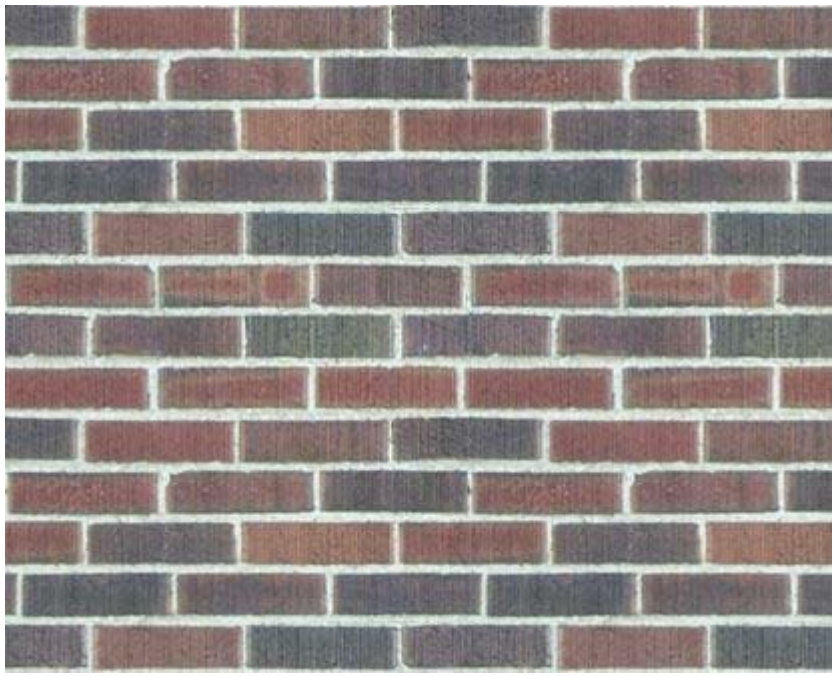


Fig. 13. Tangential section through the human fovea. Larger cones (arrows) are blue cones.

Sampling and Aliasing



Sampling and Aliasing



Photoreceptor types

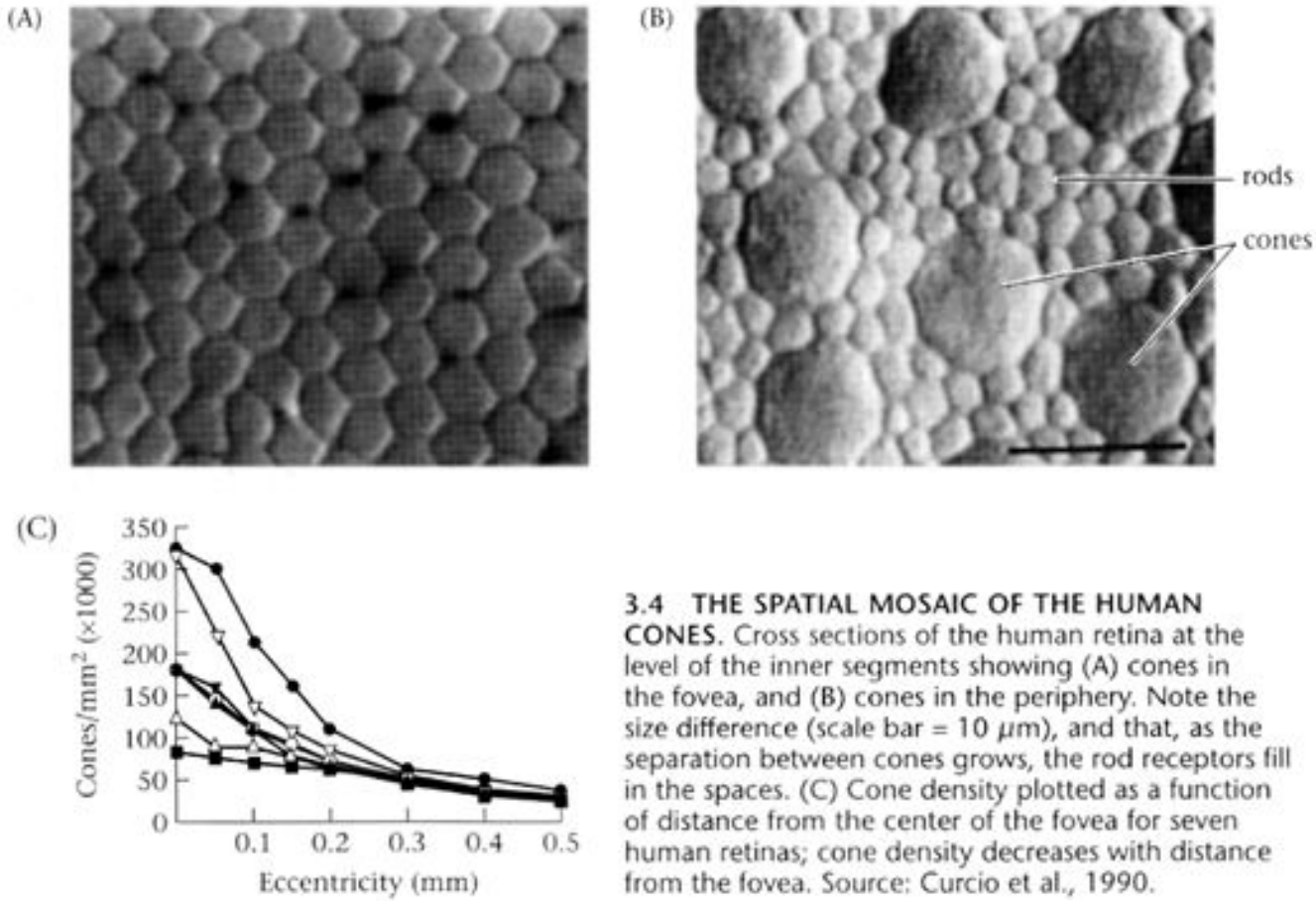
Rods

- *Scotopic* vision (low illumination)
- Do not mediate color perception
- High density in the periphery to capture many quanta
- Low spatial resolution
- Many-to-one structure
  - The information from many rods is conveyed to a single neuron in the retina
- Very sensitive light detectors
  - Reaching high quantum efficiency could be the reason behind the integration of the signals from many receptors to a single output. The price for this is a low spatial resolution
- About 10 millions

Cones

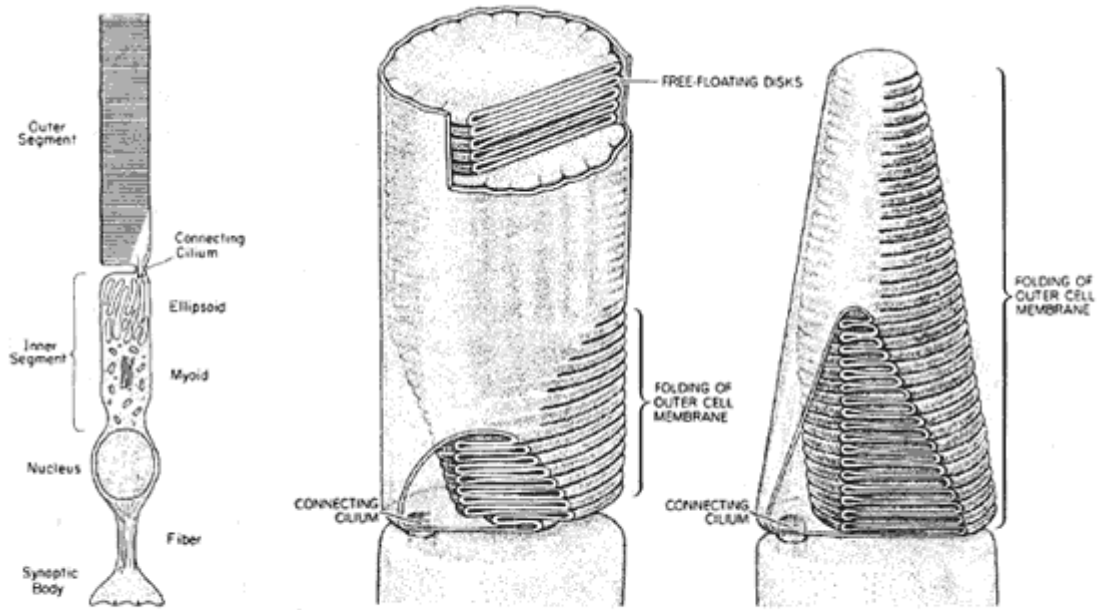
- *Photopic* vision (high illumination)
- Mediate color perception
- High density in the fovea
- One-to-one structure
  - Do not converge into a different single neuron but are communicated along private neural channels to the cortex
- High spatial resolution
  - The lower sensitivity is compensated by the high spatial resolution, providing the eye with good acuity
- About 5 millions
  - 50000 in the central fovea

Cones and Rods mosaic



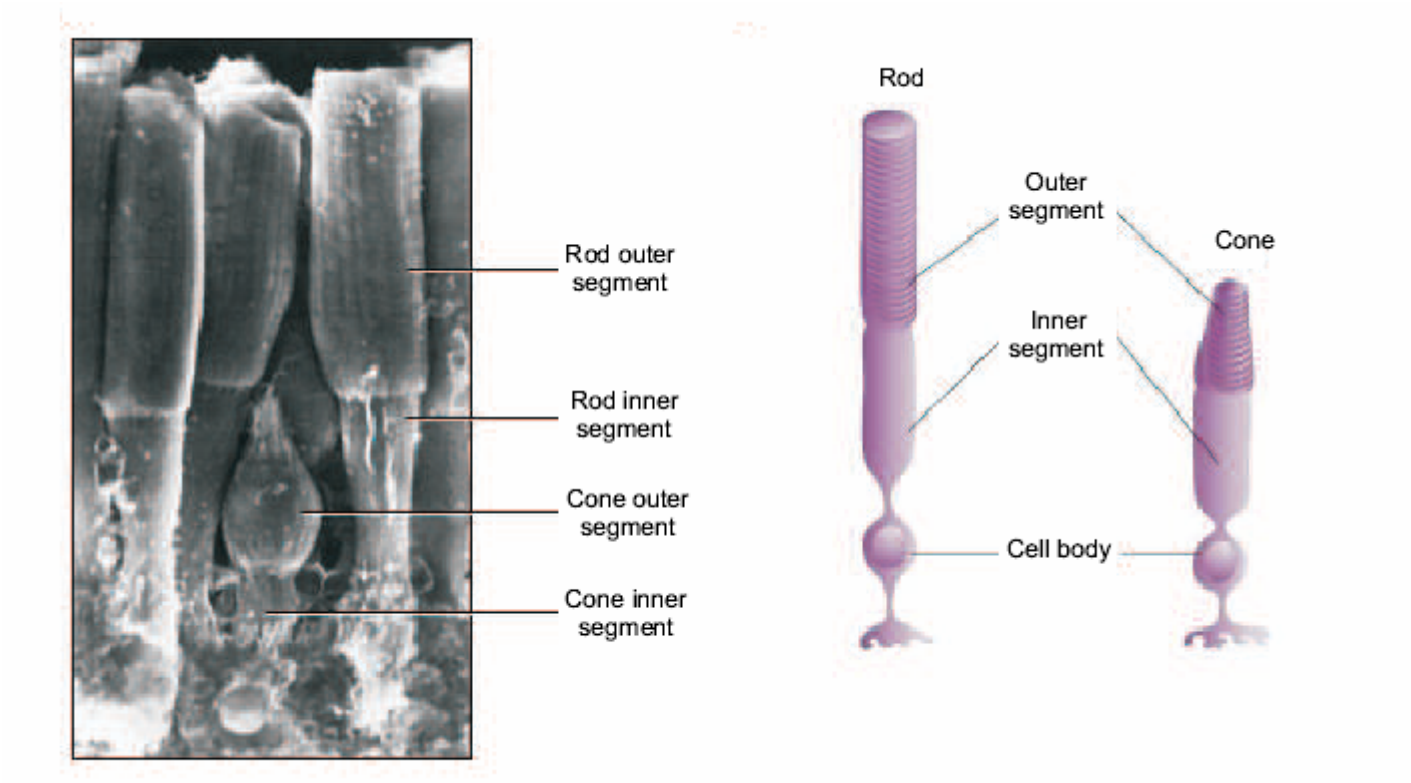
3.4 THE SPATIAL MOSAIC OF THE HUMAN CONES. Cross sections of the human retina at the level of the inner segments showing (A) cones in the fovea, and (B) cones in the periphery. Note the size difference (scale bar = 10 μm), and that, as the separation between cones grows, the rod receptors fill in the spaces. (C) Cone density plotted as a function of distance from the center of the fovea for seven human retinas; cone density decreases with distance from the fovea. Source: Curcio et al., 1990.

Cones and Rods shape



At the left is a generalized conception of the important structural features of a vertebrate photoreceptor cell. At the right are shown the differences between the structure of rod (left) and cone (right) outer segments. These diagrams are from Young (1970) and Young (1971).

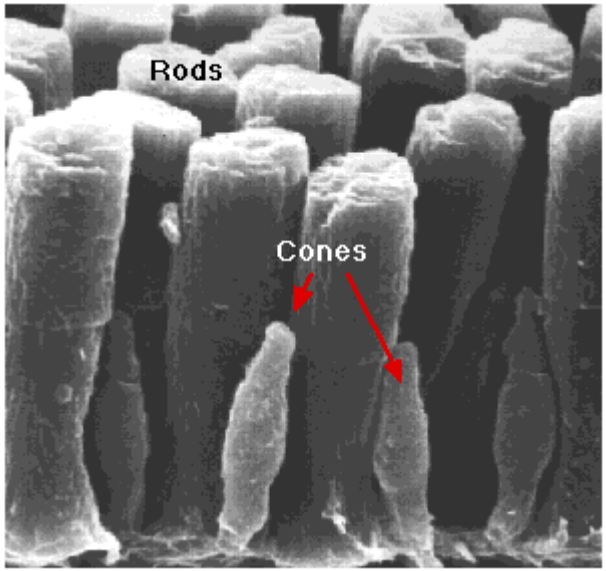
Cones and rods shapes

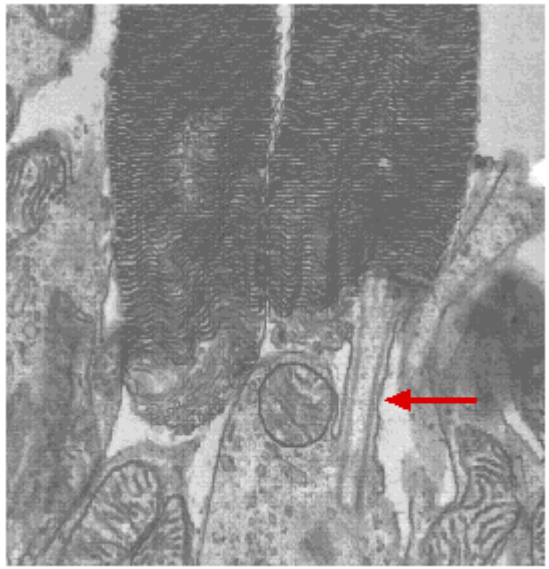


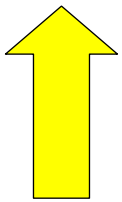
The light enters the inner segment and passes into the outer segment which contains light absorbing photopigments. Less than 10% photons are absorbed by the photopigments [Baylor, 1987].

The rods contain a photopigment called rhodopsin.

Cone and rods



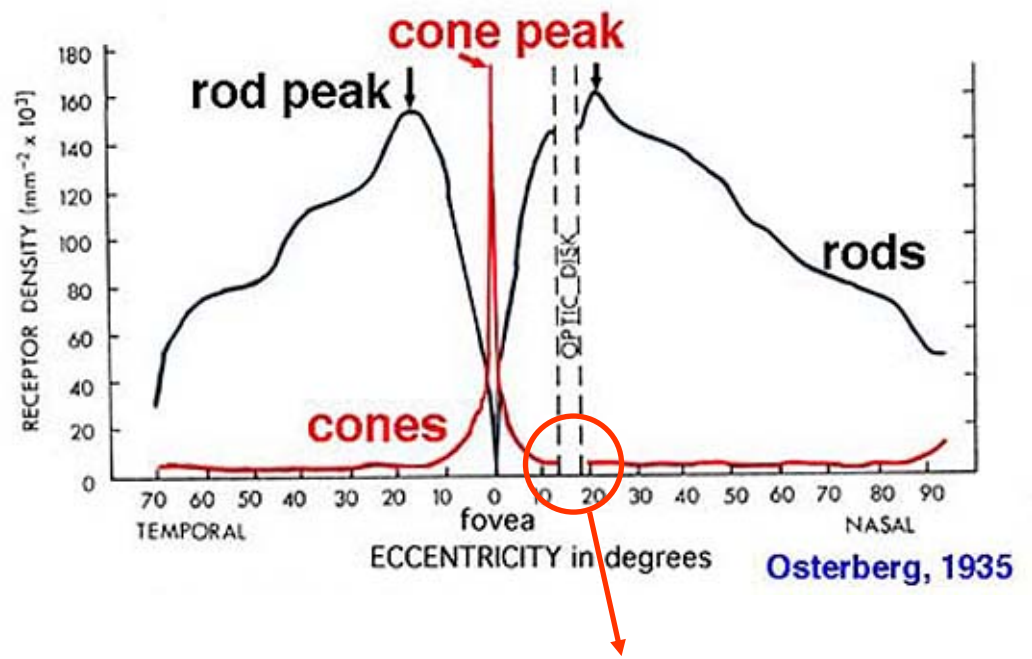




photons

The fovea

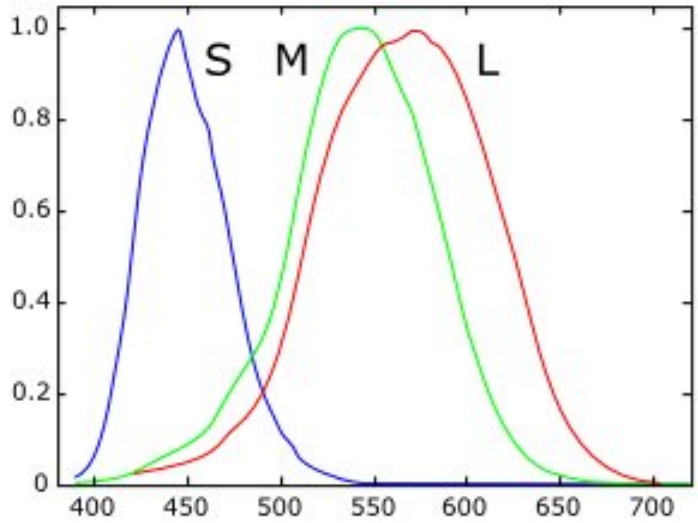
- The fovea is the region of the highest visual acuity. The central fovea contains no rods but does contain the highest concentration of cones.

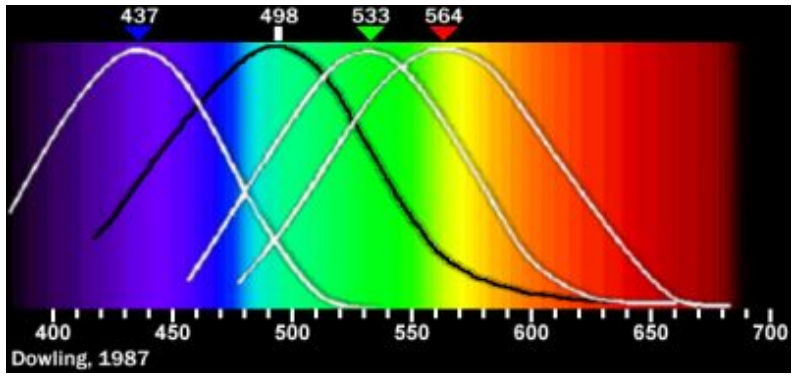


Properties of Rod and Cone Systems

Rods	Cones	Comment
More photopigment	Less photopigment	
Slow response: long integration time	Fast response: short integration time	Temporal integration
High amplification	Less amplification	Single quantum detection in rods (Hecht, Schlaer & Pirenne)
Saturating Response (by 6% bleached)	Non-saturating response (except S-cones)	The rods' response saturates when only a small amount of the pigment is bleached (the absorption of a photon by a pigment molecule is known as bleaching the pigment).
Not directionally selective	Directionally selective	Stiles-Crawford effect (see later this chapter)
Highly convergent retinal pathways	Less convergent retinal pathways	Spatial integration
High sensitivity	Lower absolute sensitivity	
Low acuity	High acuity	Results from degree of spatial integration
Achromatic: one type of pigment	Chromatic: three types of pigment	Color vision results from comparisons between cone responses

Types of cones

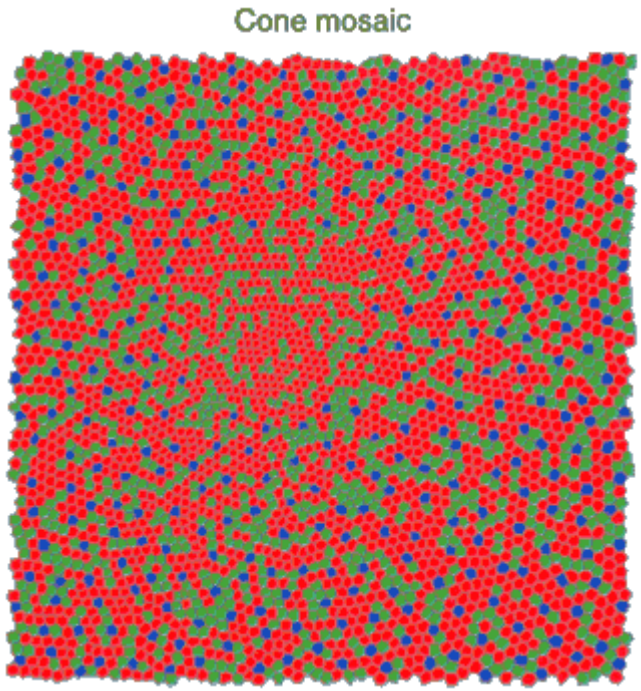




The cones are classified based on their wavelength selectivity as L (long), M (medium) and S (short) wavelength sensors.

L,M and S cones have different sensitivity and spatial distributions. The S cones are far less numerous and more sensitive than the others.

Cone mosaic



Williams (1985) measured the sampling density of the mosaic of the L- and M-cones together. His results are consistent with a sampling frequency of 60 cpd at the central fovea, consistent with a center-to-center spacing of the cones of 30 minutes of degree.

The sampling frequency than decreases when increasing the visual angle, consistently with the decrease in cone density.

This diagram was produced based on histological sections from a human eye to determine the density of the cones. The diagram represents an area of about *1° of visual angle*. The number of S-cones was set to 7% based on estimates from previous studies. The L-cone:M-cone ratio was set to 1.5. This is a reasonable number considering that recent studies have shown wide ranges of cone ratios in people with normal color vision. In the central fovea an area of approximately 0.34° is S-cone free. The S-cones are semi-regularly distributed and the M- and L-cones are randomly distributed.
Throughout the whole retina *the ratio of L- and M- cones to S-cones is about 100:1*.

Wavelength encoding

- Scotopic matching experiment → Scotopic luminosity function V'(λ)
  - Characterizes vision at low illumination conditions
  - Rod responses
  - One primary light and one test light
  - The intensity of the light beam is the parameter

- Photopic color matching experiment→ Color matching finctions (CMF), photopic luminosity function V(λ)
  - Characterizes vision under high illumination conditions
  - Cones responses
  - Three primary lights and one test light
  - The intensities of each primary lights are the parameters

Representation

Representation

Visual streams

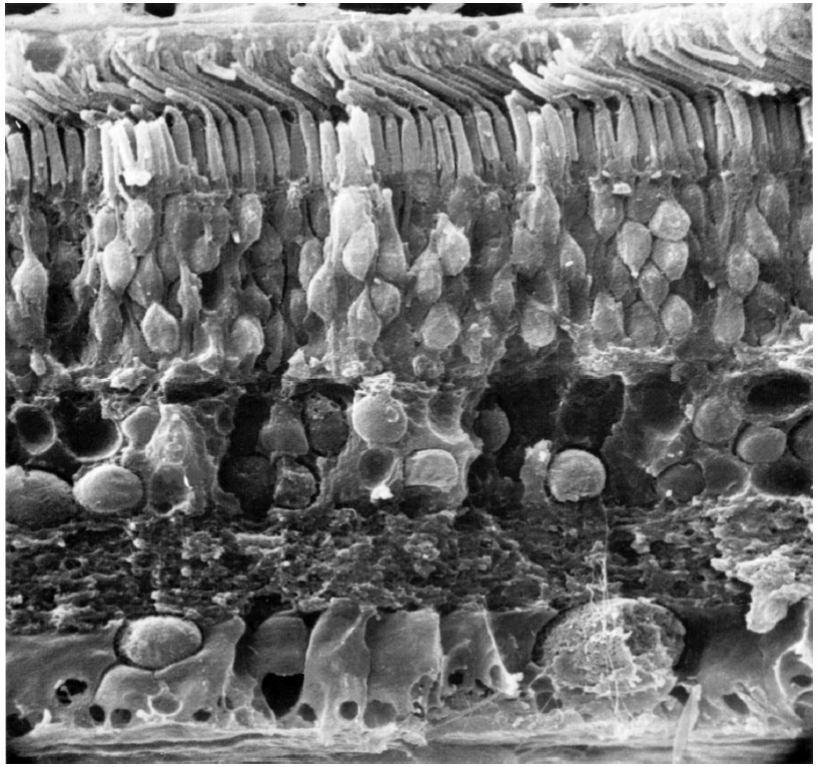
- The visual system consists of a collection of pathways, each responsible for analyzing different aspects of the retinal image.
- It begins at the early stages of visual encoding, with the segregation between rods and cones. The *specialization* is elaborated in the retina and continues into the **cortical area**.

Issues of interest

- Adaptation and contrast
  - Compensation in response to variations of the illumination level (adaptation)
- Multiresolution
  - Image contrast is represented *within a certain range of spatial frequencies and orientations*, as shown by both behavioral and electrophysiological studies
- Linking hypothesis
  - About matching behavioral and biological measurements
  - Example: linking color matching experiments with cone sensitivities

The Retinal representation

Retina cross section (rabbit)



The human retina consists of different layers and its structure changes with eccentricity.

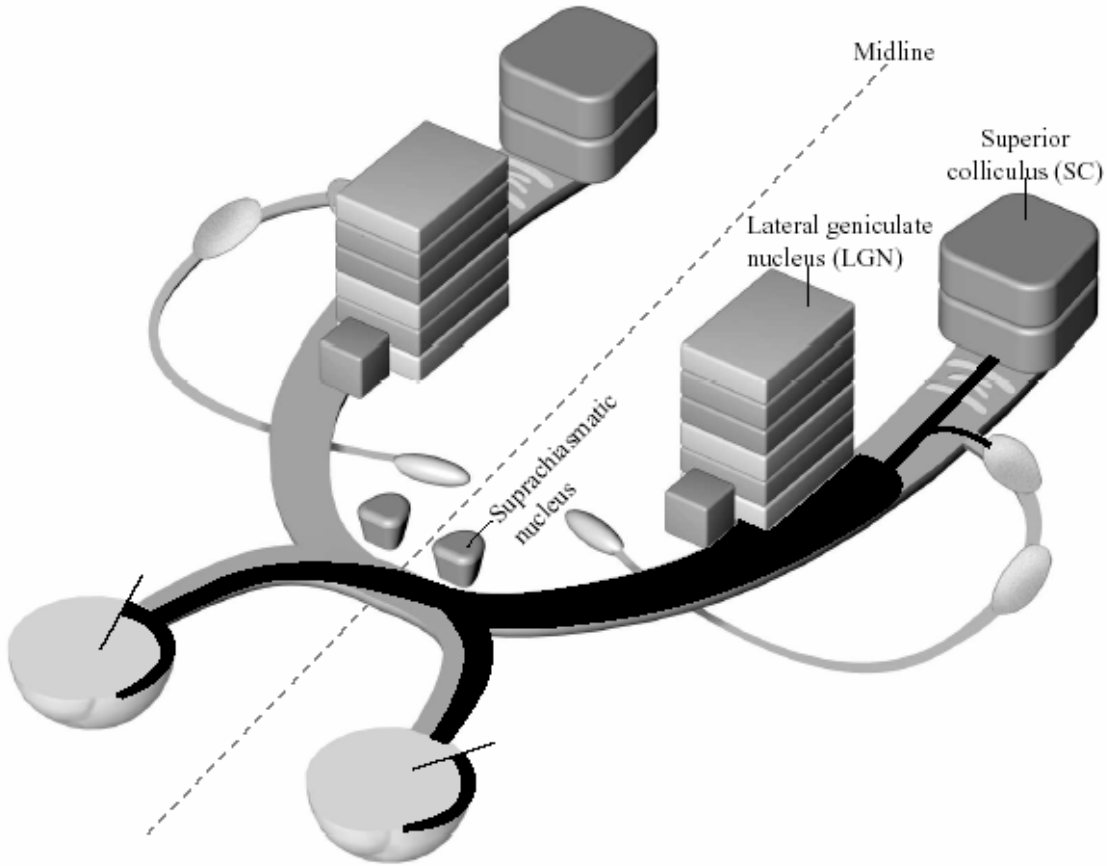
There are many types of cells and of interconnections both within and between layers, which make the system quite complex.

The connected series of neurons carrying information **in parallel** are referred to as visual pathways or **visual streams**. It seems that the segregation to different pathways starts at the output of rods and cones.

One of the main functions of the retina is to organize the information collected by the photoreceptors into a collection of visual streams.

It is common belief that each visual stream carries an efficient representation of the spatiotemporal component of the image that is *most relevant for tasks* carried out in the visual area where the ganglion cell output is sent.

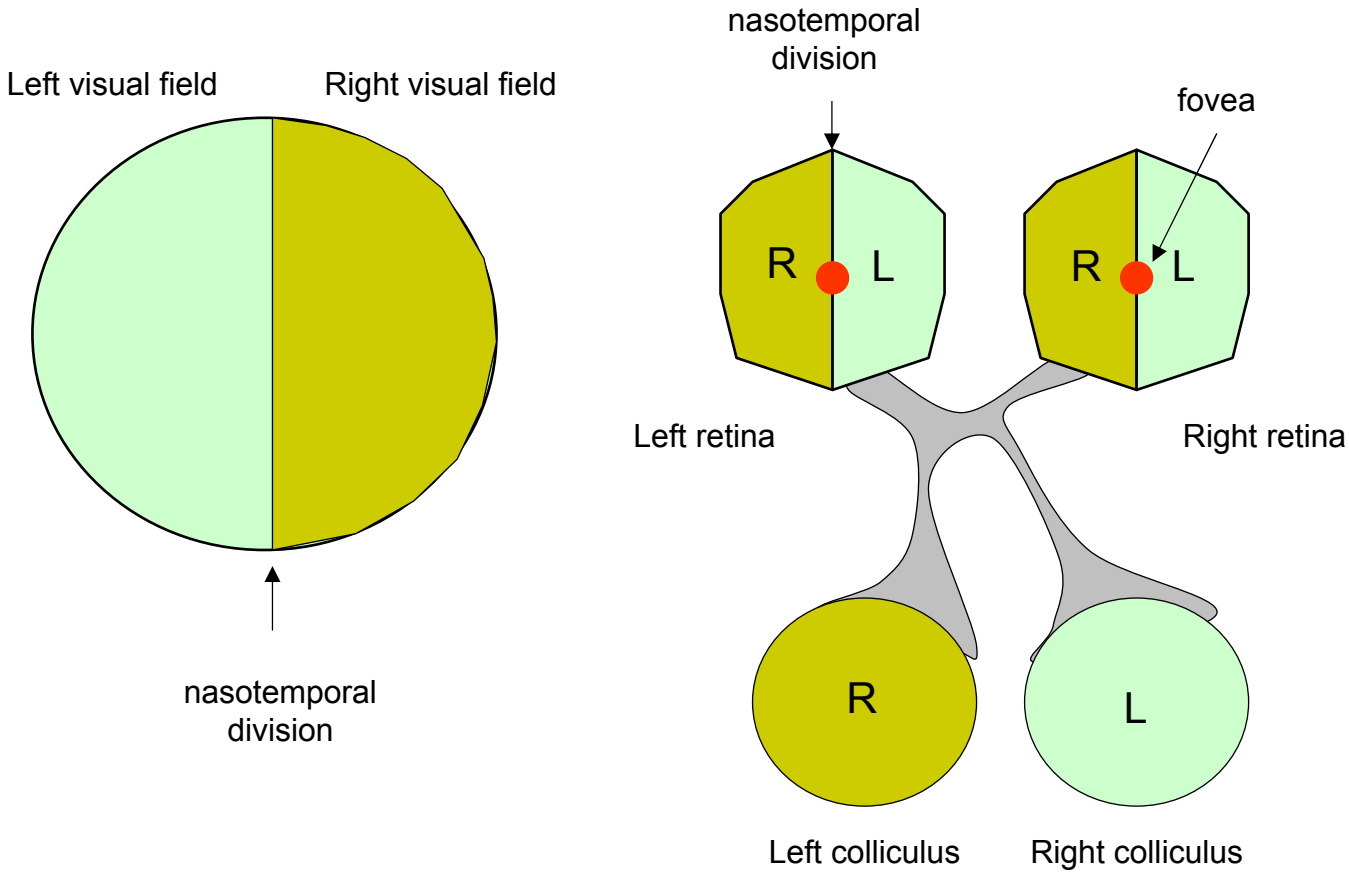
Neural information flow



Along the pathway

- Lateral Geniculate Nucleus (LGN)
  - The LGN is placed *between the retina and the cortex*. It resembles a six-layered, warped cake.
    - The lower two layers contain large cell bodies, called *magnocellular* neurons, while the upper four layers are characterized by small cell bodies termed *parvocellular* neurons.
  - About 90% of the fibers in the optic nerve project to the lateral geniculate nucleus (LGN) of the thalamus and from there onto primary visual cortex.
    - This pathway dominates conscious visual perception. About 100,000 ganglion cells project to the superior colliculus (SC) at the top of the midbrain.
- Superior Culliculus (SC)
  - The superior colliculi retains a number of important visual functions underlying orienting responses as well as eye and head movements.
  - The SC integrates visual and auditory information together with head motion and directs eyes to regions of interest in the external world (saccades).
    - The two SCs are the most important visual regions in fish, amphibians, and reptiles. In primates, much of their function has been taken over and extended by the cortex.
    - The SC can be divided conveniently into superficial, intermediate, and deep layers. The superficial layer receives direct input from retinal ganglion cells in a topographic manner.
    - Each nucleus is about 6mm wide.

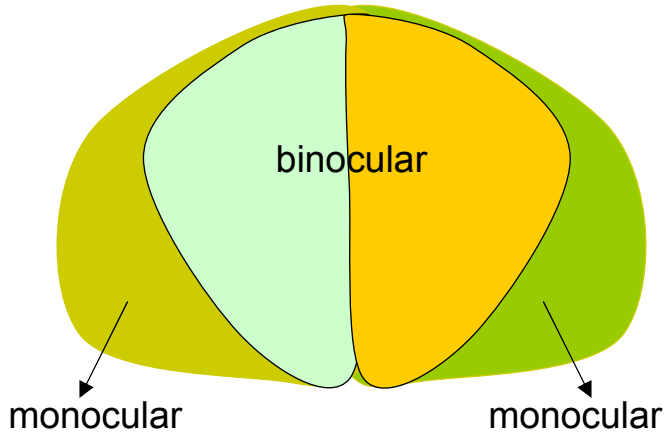
Symmetries



The view is **from behind**, looking forward at the backs of the retinas.

Both retinas send the information **from the right visual field to the left nucleus**, and viceversa. Each nucleus receives from the portions of both retinas that receives the image of the controlateral visual field

Monocular and binocular vision



The *temporal* retina of the *right* eye views the *left* visual hemifield, but its view is limited because of the nose.

The *nasal* retina of the *left* eye has a more extensive view of the visual field. The portion of the visual field that only this eye sees is called the ***monocular* portion** of the left visual field.

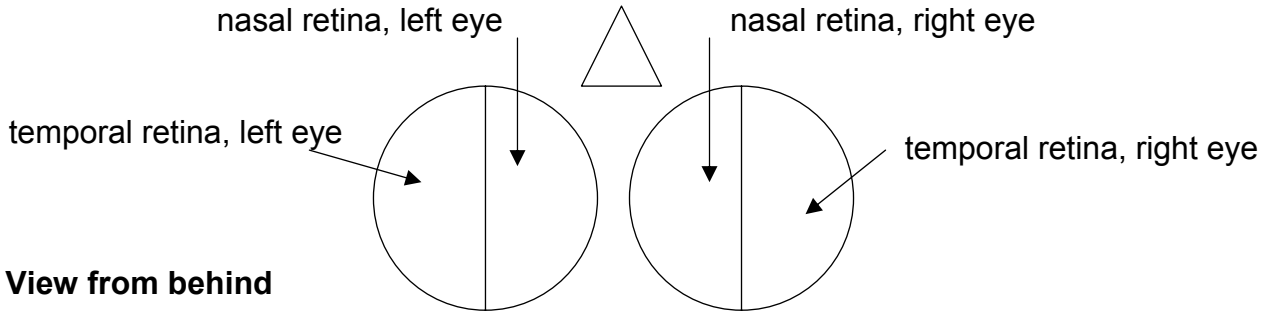
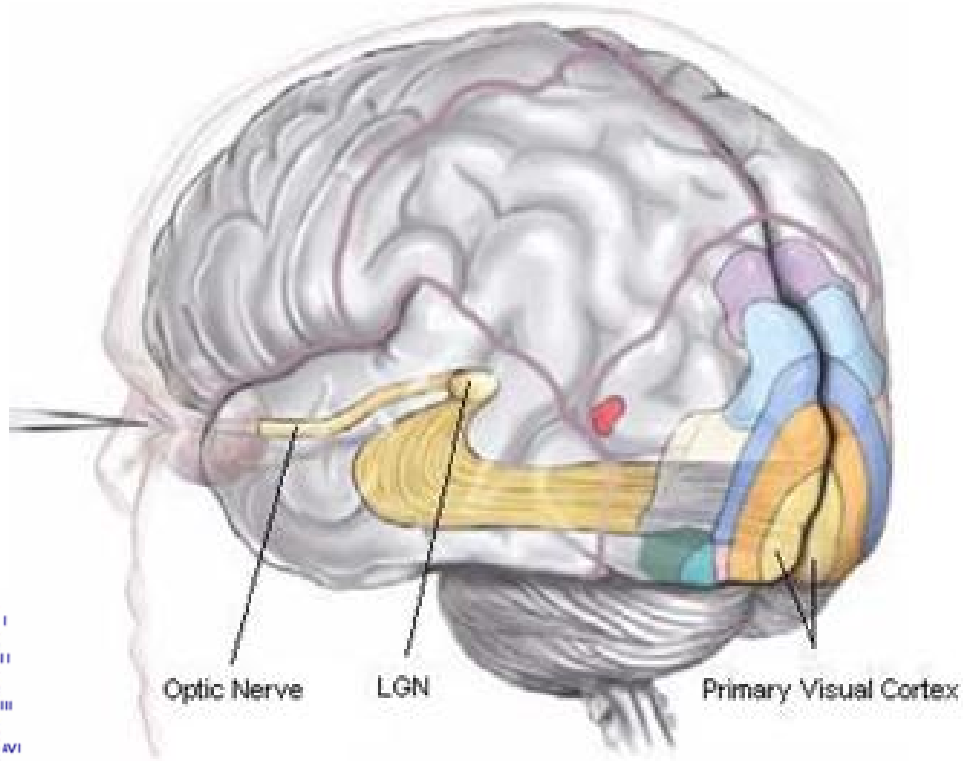
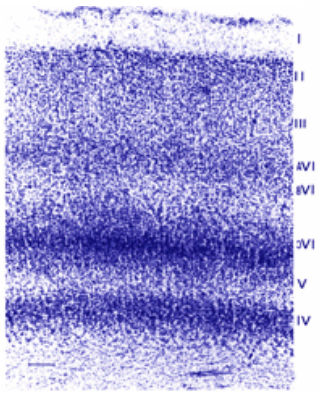


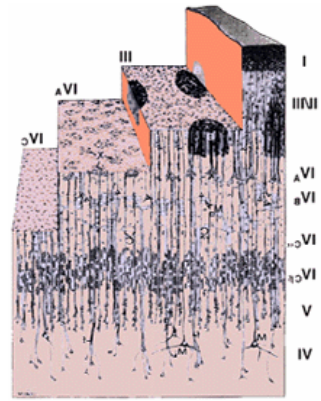
Image contrast and adaptation

- Challenge: The neurons must be sensitive to image patterns despite the great variation of illumination along the day
  - The neurons response range covers about 2 or 3 orders of magnitude, versus the 6 of the daylight illumination
- Solution: Encode the *local contrast* instead of the absolute stimulus values
  - Local contrast: percentage change in the image intensity relative to local average
    - remains constant despite the changes in the illumination

Cortical representation







Feedforward and Feedback Pathways

- The signals from the two retinas are communicated to area V1 via the LGN. In the macaque monkey, after the signals are processed in V1 they are communicated via multiple pathways to the 30+ visually responsive extrastriate cortical areas
- Anatomical hierarchical models place the visual cortical areas into a multi-level processing model based upon the pattern of *feed forward, lateral, and feedback pathways* found in each area.
- It is true that visual information is passed from the retina to the LGN to V1 to the higher cortical areas and that the higher areas project back to V1 and V1 to the LGN

➢ Very complex path of interaction

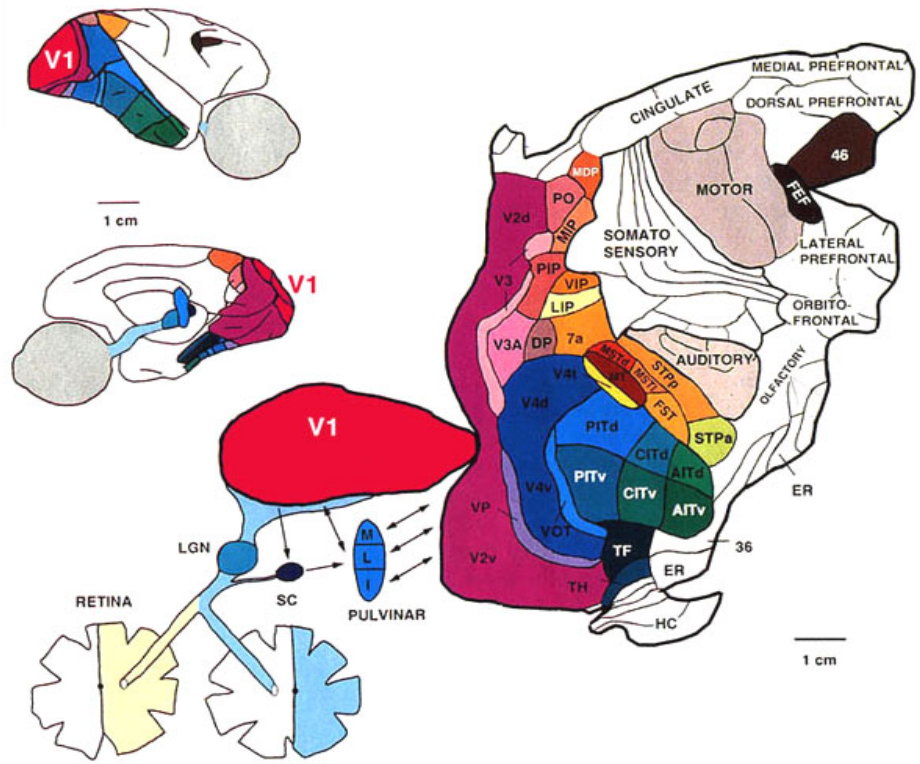
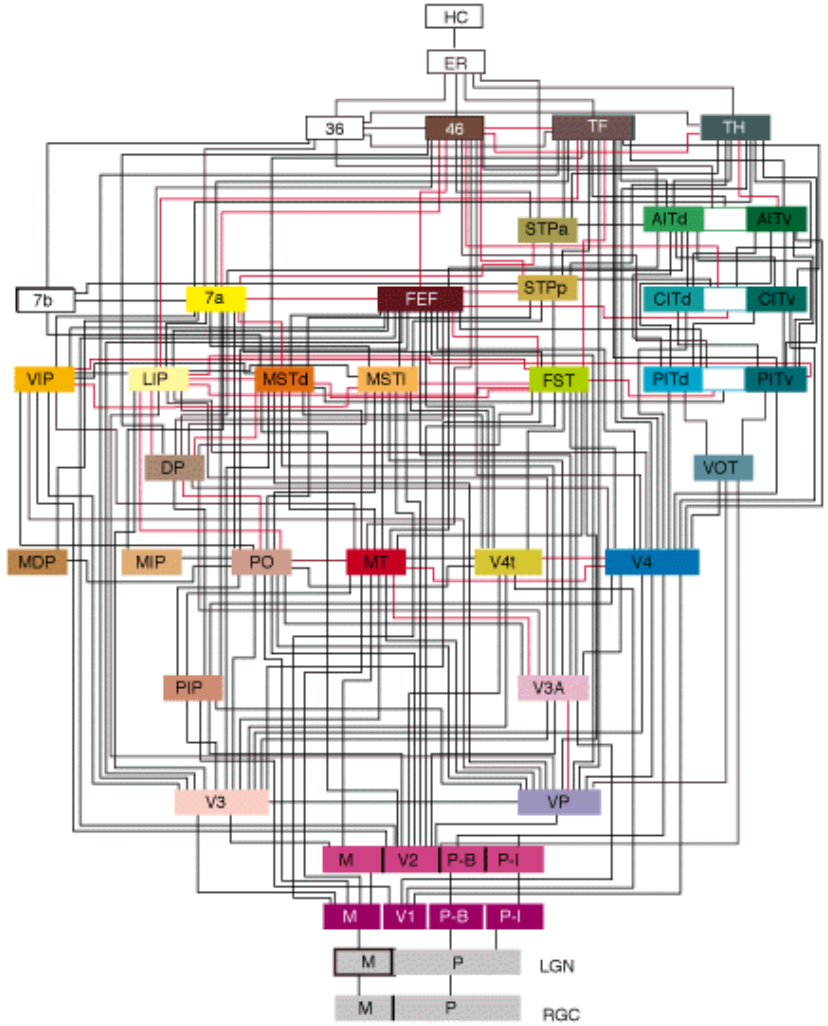


Figure 19. Much of V1 is located in the calcarine sulci and its relationship to other brain areas is best shown by unfolding the brain and showing it flattened open. The visually responsive areas of the macaque monkey are shown in color. From Van Essen et al. (1992).

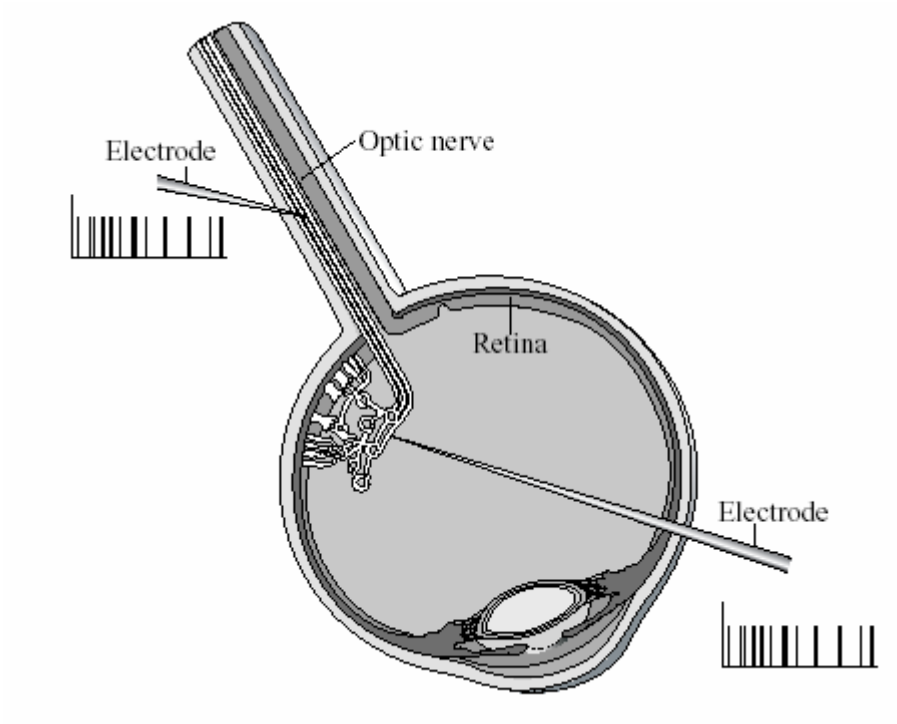
Visual area wiring diagram



Receptive fields

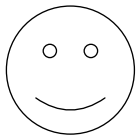
Retinal ganglion cells response to light

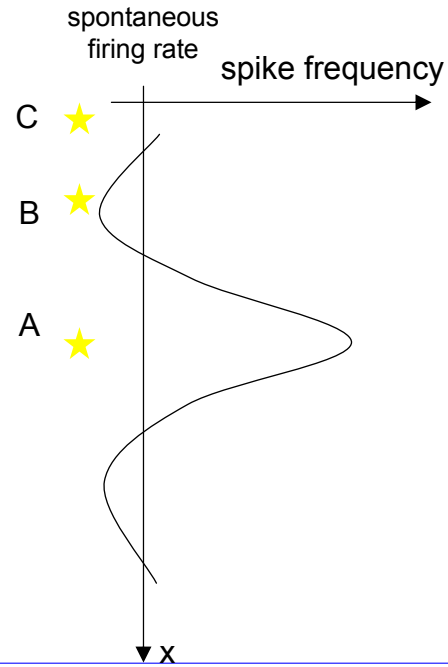
- Action potentials or spikes
  - The responses of ganglion cells are analyzed by recording the *temporal pattern* of action potentials caused by light stimulation
  - These can be measured by either placing a microelectrode near to their cell bodies in the retina or in the optic nerve outside the eye
- Receptive field [Kuffler, 1952]
  - Operationally, the receptive field can be defined as the *portion of the visual field in which an appropriate stimulus modulates the cell response*
  - The *RF depends on the entire visual pathway*, though there is no feedback in the retinal ganglion cells



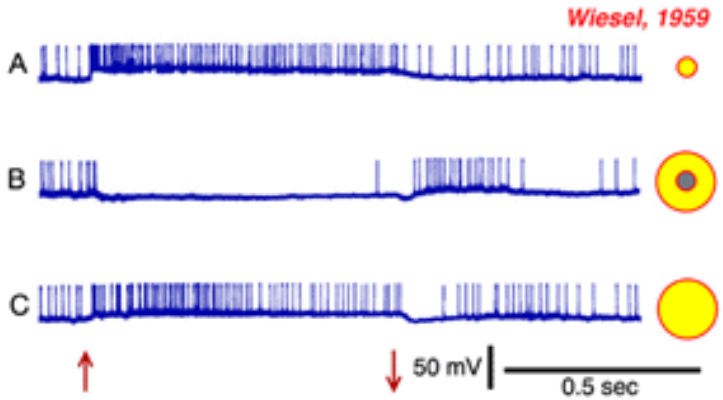
Center-surround organization

- Spontaneous firing rate
  - Average number of spikes per unit time in presence of a constant field
    - Typically 50 spikes/sec
- The RF is characterized through the *change* in firing rate caused by a stimulus at different positions in the visual field



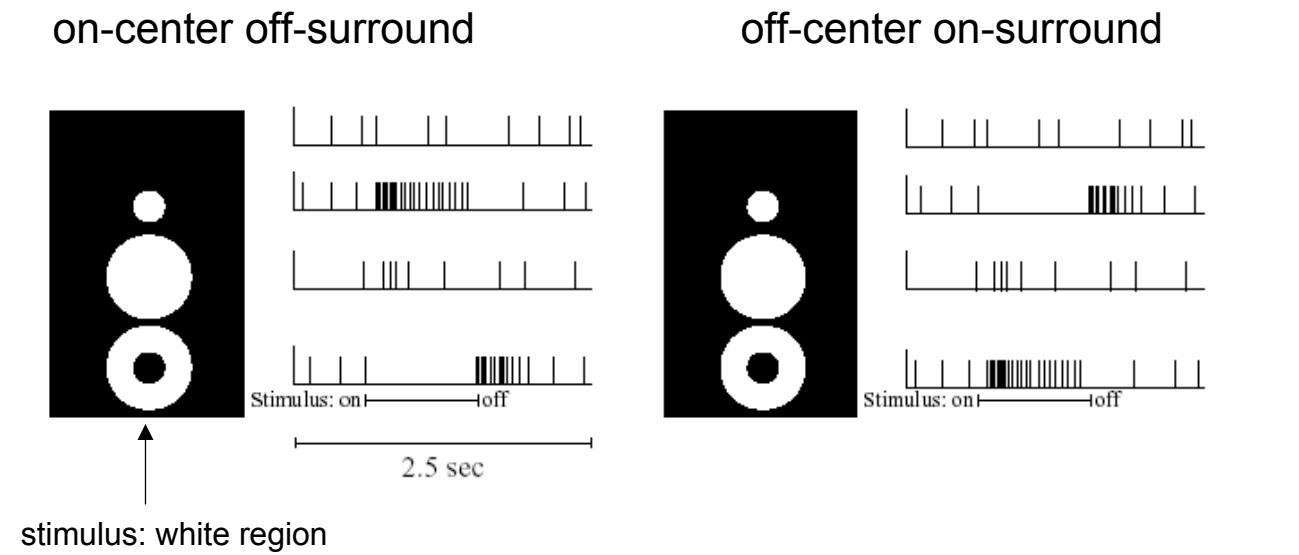


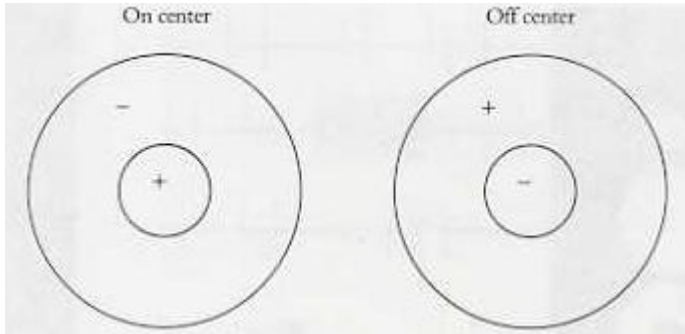
on-center off-surround



Center-surround organization

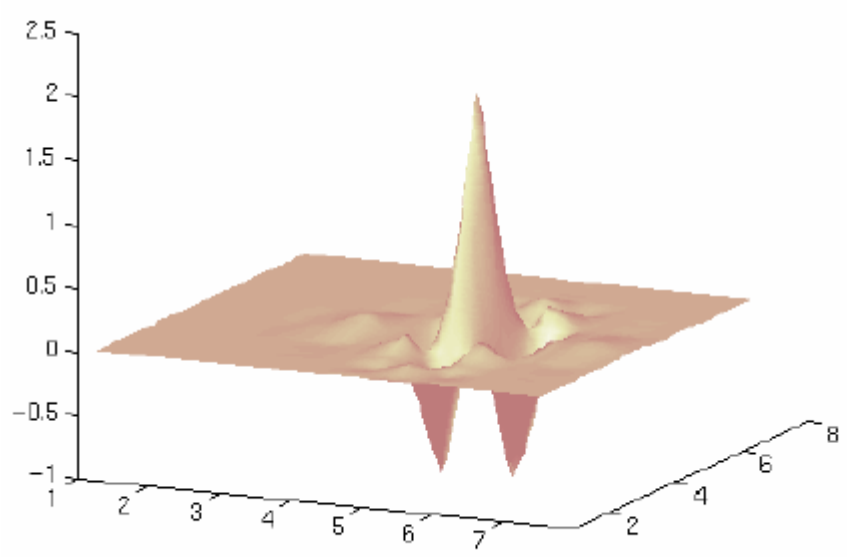
Left: Four recordings from a typical on-center retinal ganglion cell. Each record is a single sweep of the oscilloscope, whose duration is 2.5 seconds. For a sweep this slow, the rising and falling phases of the impulse coalesce so that each spike appears as a vertical line. To the left the stimuli are shown. In the resting state at the top, there is no stimulus: firing is slow and more or less random. The lower three records show responses to a small (optimum size) spot, a large spot covering the receptive-field center and surround, and a ring covering the surround only. Right: Responses of an off-center retinal ganglion cell to the same set of stimuli shown at the left.





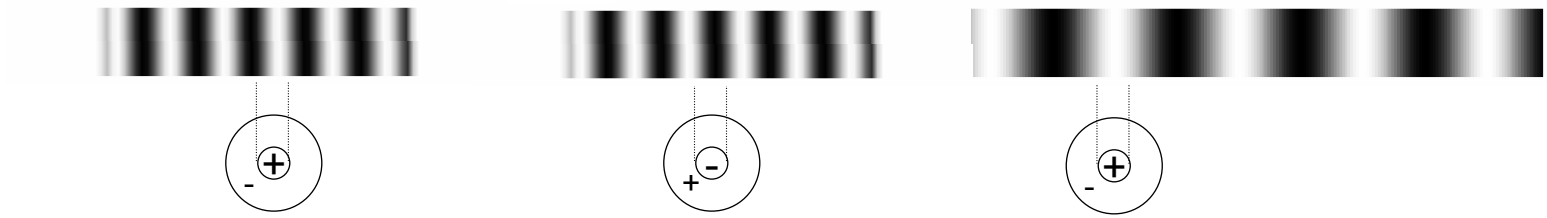
Two-dimensional RF

- The shape of the RF reflects many properties of the neurons, like the sensitivity to different frequencies (bandwidth)
- Linearity holds, so one can characterize the response to complex stimuli by exploiting homogeneity and superposition
  - Test for linearity in retinal ganglion cells [Enroth, Cugell and Pinto, 1970]



Contrast Sensitivity Functions

- Contrast threshold
  - Amount of stimulus contrast needed to elicit a criterion level of response from the neuron
  - Principle
    - When the stimulus pattern is matched with the RF of the neuron, a small amount of contrast will elicit the criterion response level
- Contrast *sensitivity* = 1/Contrast Threshold
- The *highest spatial frequency* to which the cell responds is bounded by the size of the RF center

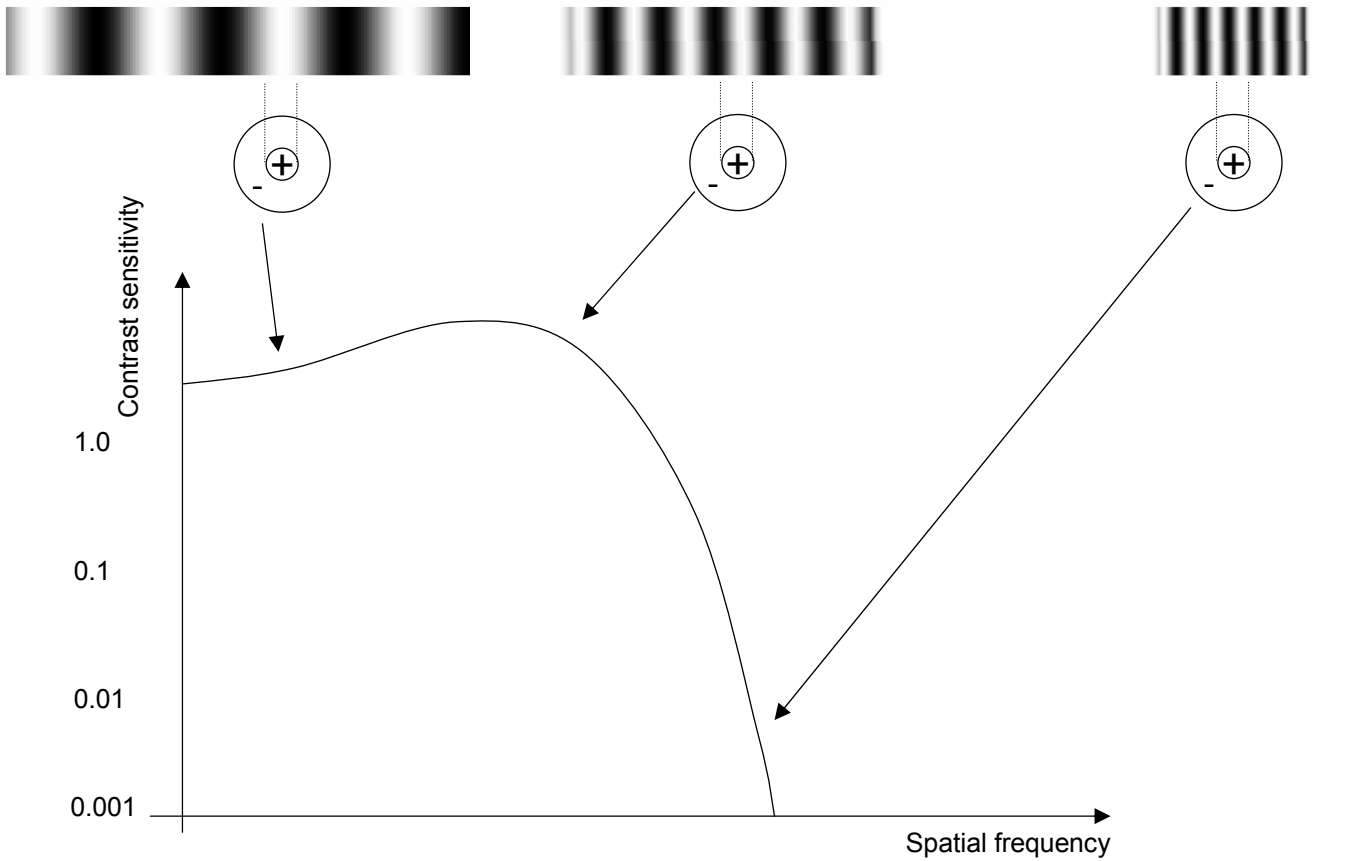


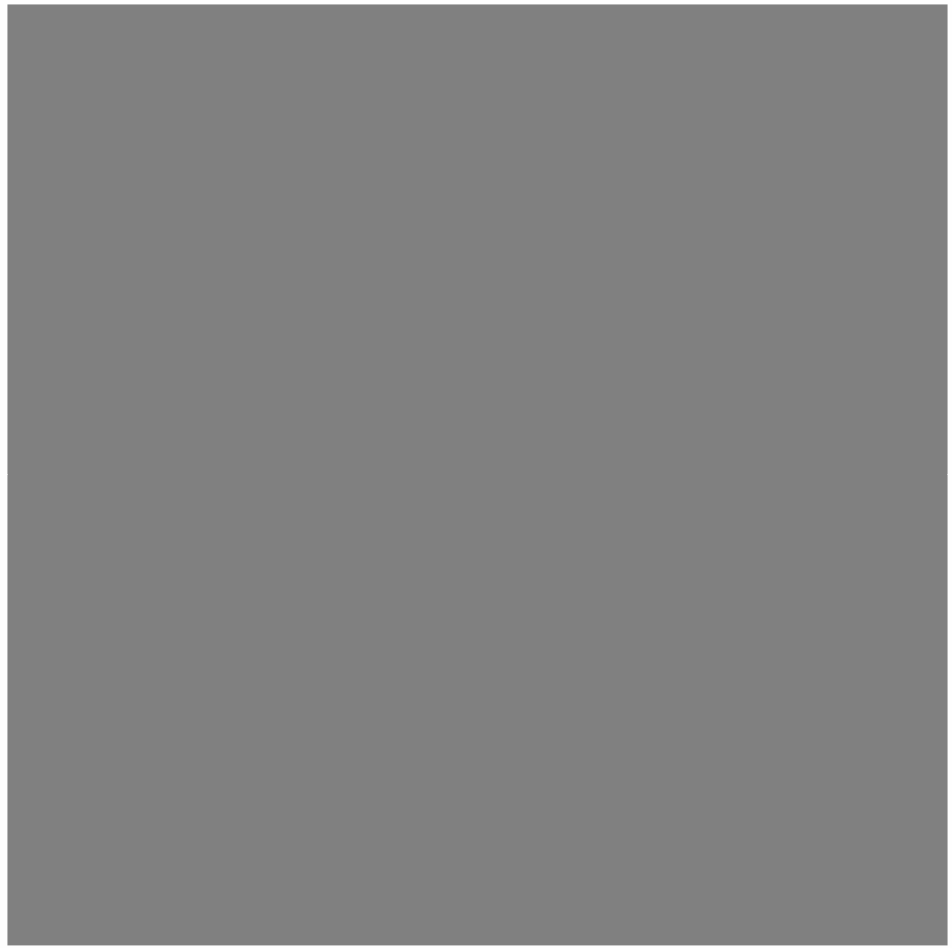
high response → low contrast threshold (on-center)

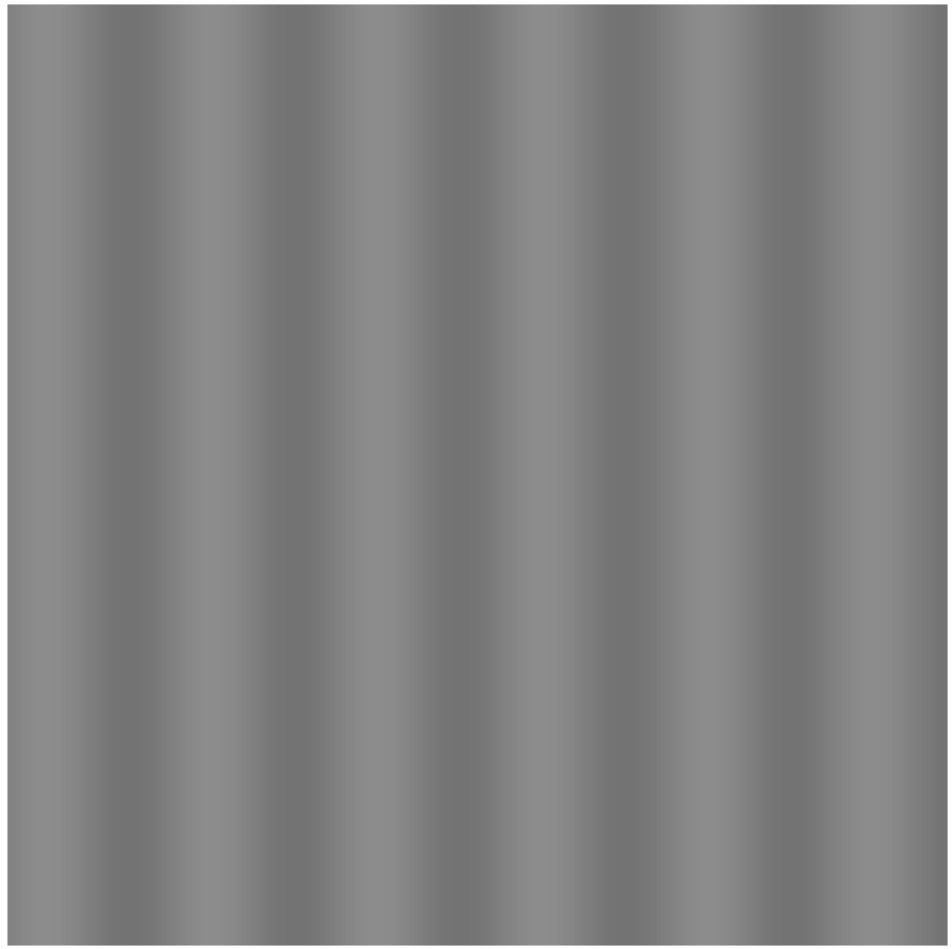
high response → low contrast threshold (off-center)

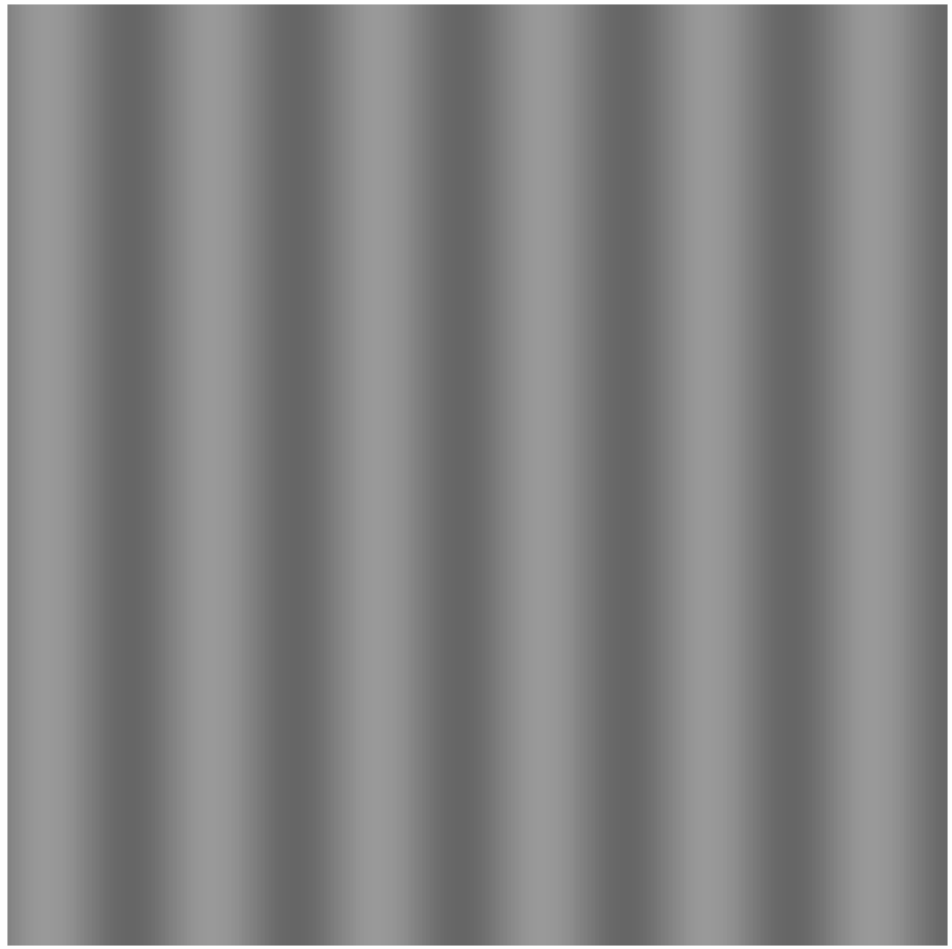
low response → high contrast threshold

Contrast Sensitivity Functions



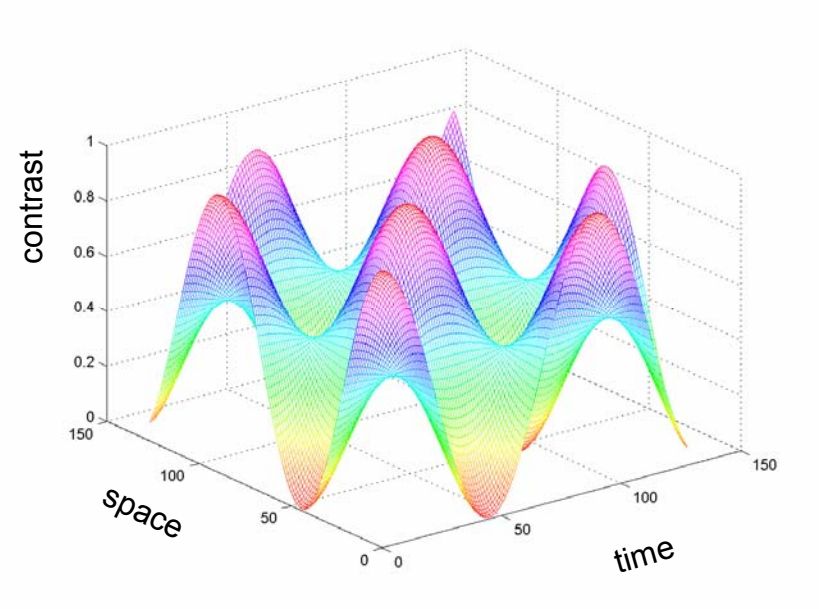


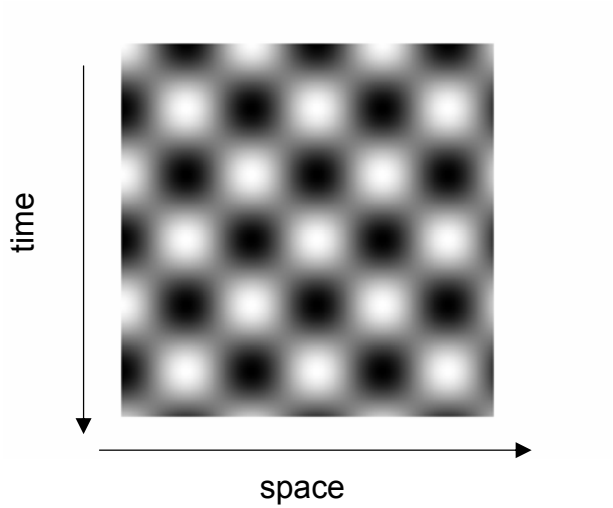




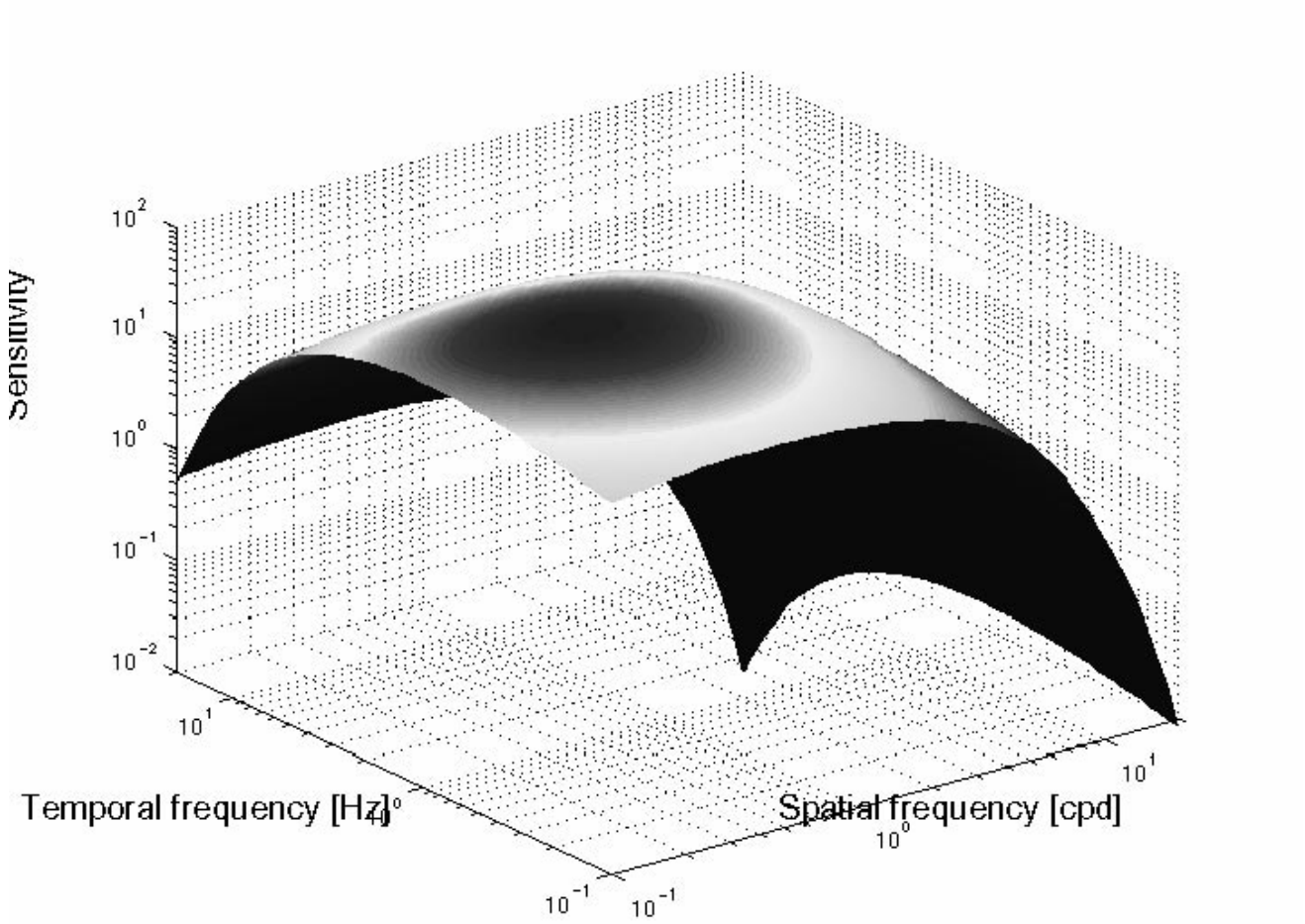
Spatio-temporal CS

- Flickering stimuli
  - The temporal frequency of the flicker f_t is an additional control variable
  - For a single neuron, the responses to many repetitions of the stimulus must be collected and averaged → *peri-stimulus time histogram* (PSTH)
- Space-time receptive fields





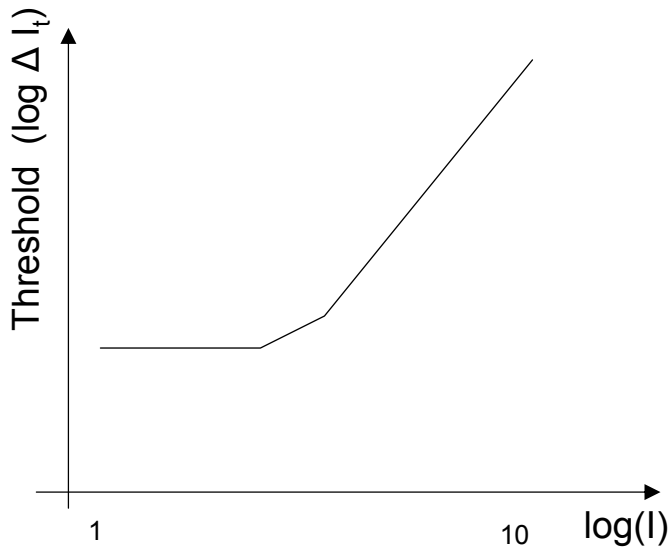
Spatio-temporal CSF



Space time separability does not hold!

Weber's law

- Threshold sensitivity as a function of background intensity



$$\frac{\Delta I_t}{I} = k$$ at threshold

I : steady value of the background intensity
ΔI_t : incremental threshold

The incremental threshold is proportional to the absolute value of the mean background intensity → *contrast sensitivity is constant* (ΔI_t /I represents the contrast at threshold)

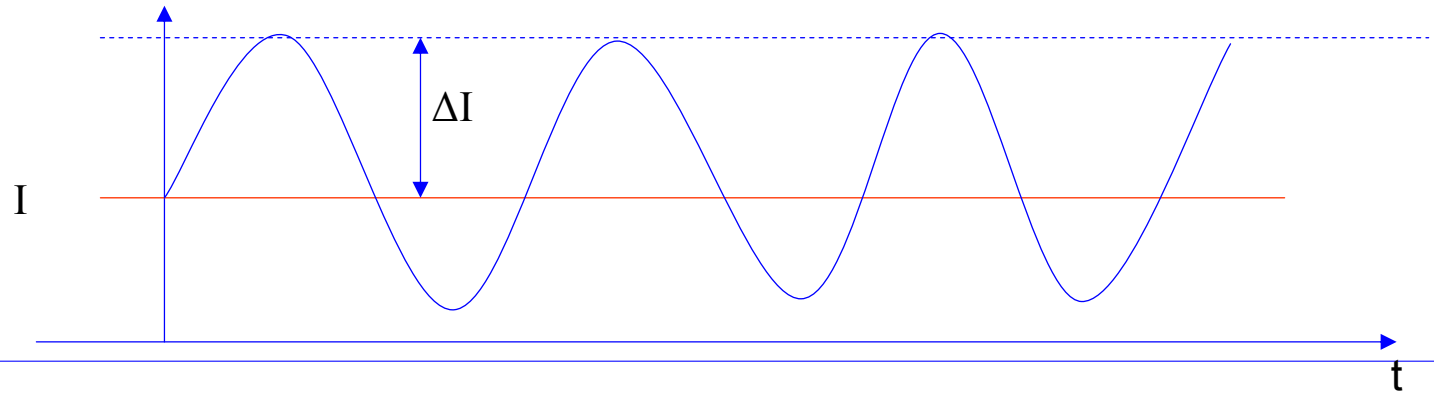
Weber's law characterizes many visual mechanisms, and is one the most important laws in vision. It is an approximation and it applies best to low spatial frequency patterns.

Good agreement with behavioral CSFs [Pasternak&Merigan, 1981, cat]

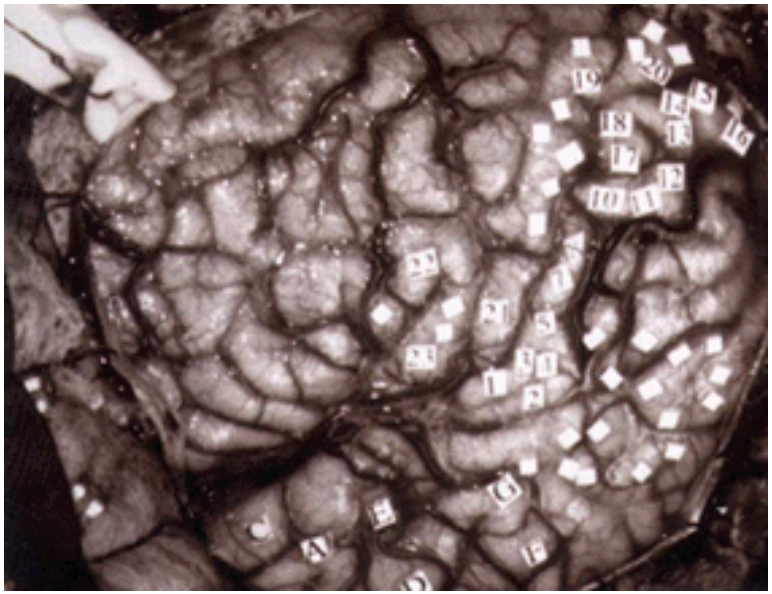
Weber's law

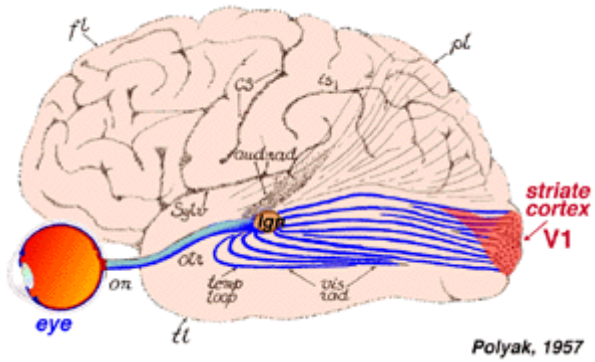
- **Weber's Law** states that the ratio of the **increment threshold** to the background intensity is a constant.
  - So when you are in a noisy environment you must shout to be heard while a whisper works in a quiet room.
  - And when you measure **increment thresholds on various intensity backgrounds, the thresholds increase in proportion to the background**.

  - The fraction **ΔI/I** is known as the **Weber fraction**. If we rearrange the equation to ΔI=IxK, you can see that *Weber's Law predicts a linear relationship between the increment threshold and the background intensity*.



The Cortical Representation

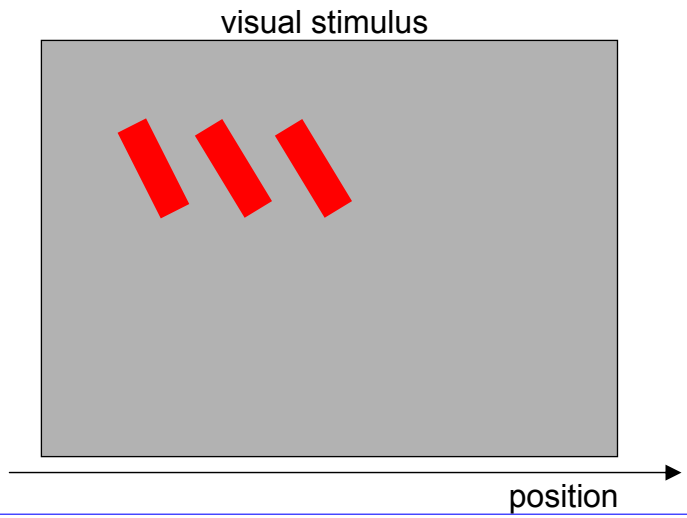


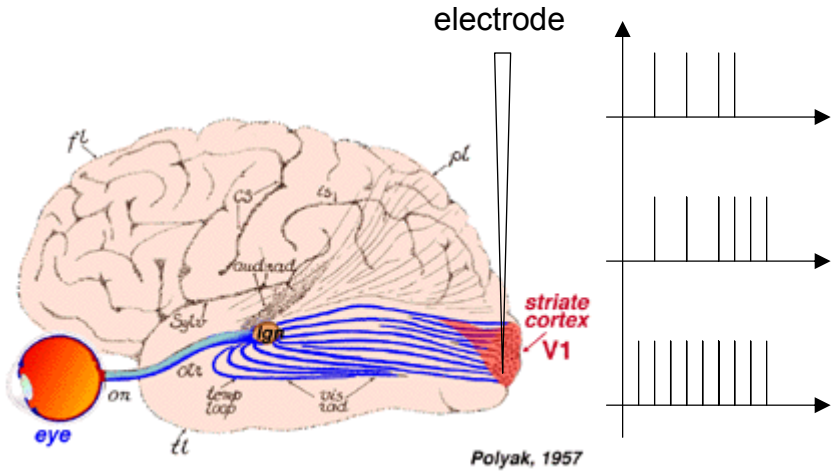


The human cortex is a 2mm thick sheet of neurons with a surface area of 1.400 cm^2, in the form of a crumpled sheet stuffed into the skull.
In primates the great part of the signals from the retina and LGN arrives at a single area called V1, or **primary visual cortex**. It comprises about $1.5x10^8$ neurons.
More than 20 other areas have been discovered to receive visual inputs.

Receptive fields in V1

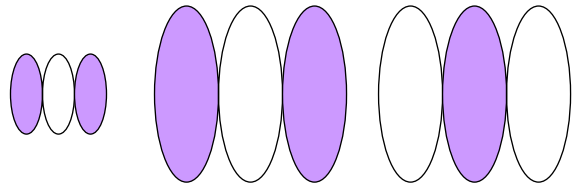
- The RF in V1 are qualitatively different from those of LGN
  - The LGN neurons' RF are circularly symmetric while the V1 are not
  - Direction selectivity
  - Orientation selectivity
  - Some of them are binocular
  - [Hubel&Wiesel, 1959-82]: **simple** cells and **complex** cells. While simple cells succeeded the test for linearity, complex cells did not
  - The classic method for testing orientation and direction selectivity is to measure the spike rate of a single cell in response to drifting oriented luminance bars and/or drifting luminance spots





Orientation selectivity

- Orientation selectivity is modeled by a RF that is elongated along the preferred orientation of the neuron
  - Stimuli oriented along the main axis of these RF are more effective at exciting or inhibiting the cell than stimuli in other orientations
  - The degree of orientation selectivity is represented by the size of the RF: the neuron with a longer RF will respond well to a narrower set of orientations
  - Orientation selectivity could result from pooling the responses of sets of neurons with symmetric RF according to different policies. The exact mechanisms underlying orientation selectivity are still mostly unknown
  - The preferred orientation of neurons varies in an orderly way that depends on the position in the cortical sheet



  - Open issues
    - What are the rules for making the interconnections that lead to the spatial organization of orientation selectivity?
    - What functional role do they have in perceptual processing?
    - Is the spatial organization essential?

Orientation selectivity

Histograms of cells' responses as a function of time

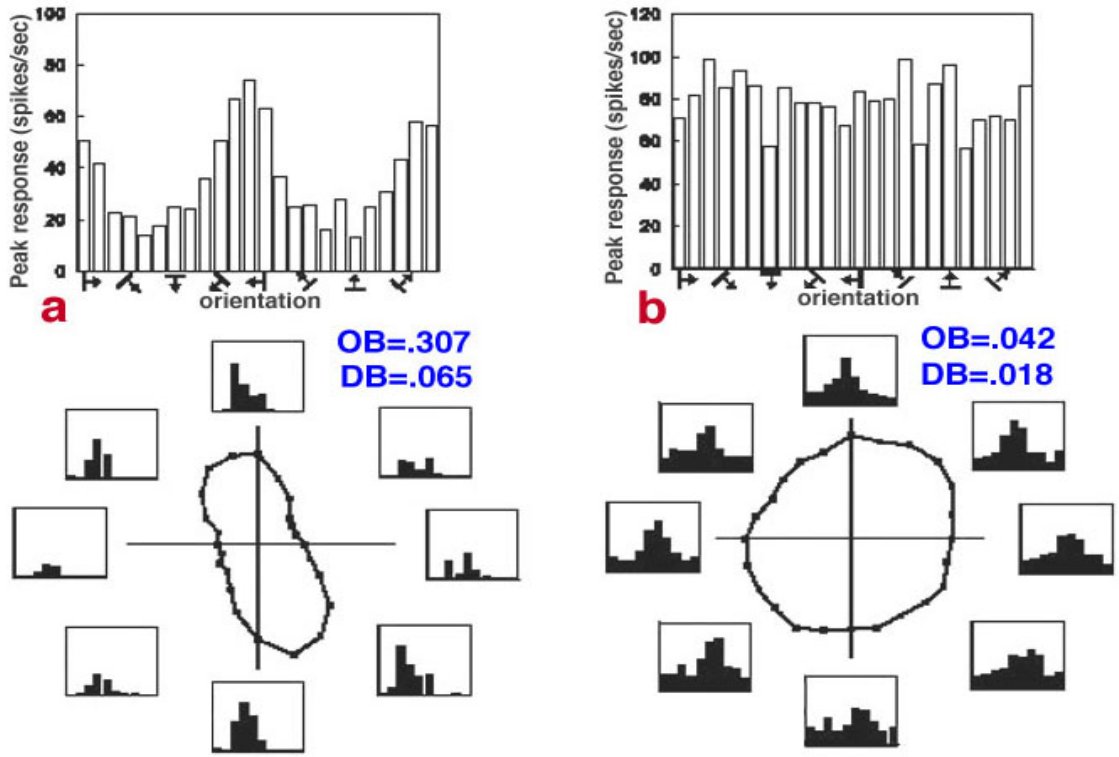
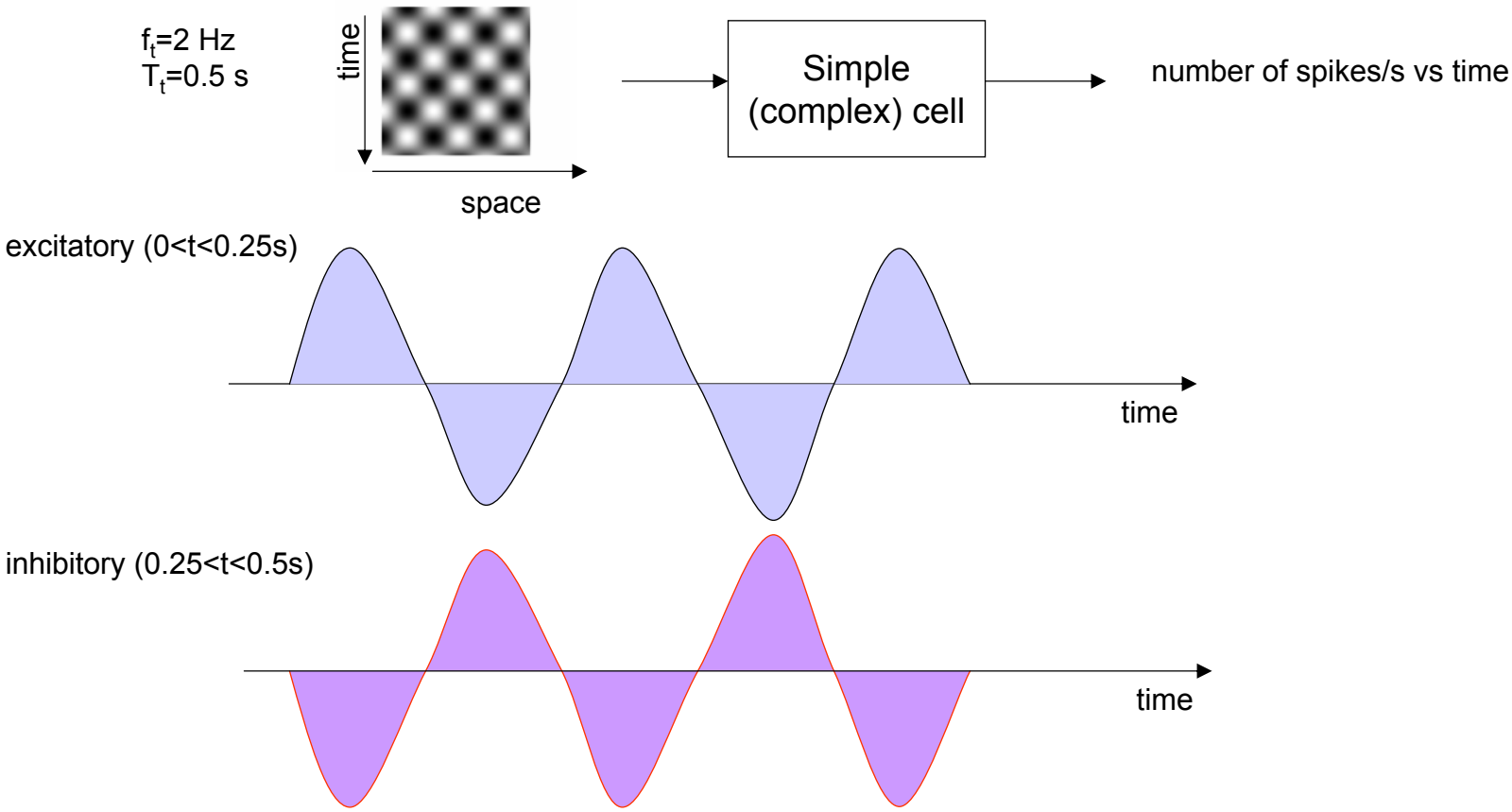


Figure 21. A tuning curve and corresponding polar plot obtained from two macaque V1 cells in response to drifting luminance bars of systematically varied orientation and direction. The responses of one orientation selective cell and one nonselective cell are provided for comparison. Histograms surrounding the polar plots demonstrate the cellular response as a function of time. Orientation bias (OB) and direction bias (DB) are measures of how selective a cell is, where >0.1 is significant, and 0.3 is approximately an 8:1 maximum firing rate to minimum firing rate ratio. From Schmolesky et al. (2000).

Non-linear features

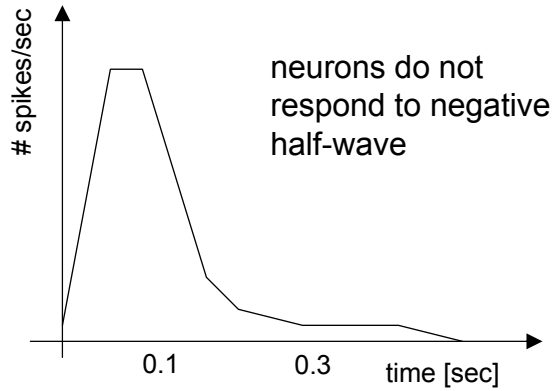
- Contrast sensitivity of simple and complex cells
  - Contrast normalization
  - Model for contrast-gain control [Heeger 1992, Simoncelli&Shwartz 2000]



Non-linear responses

Simple cells

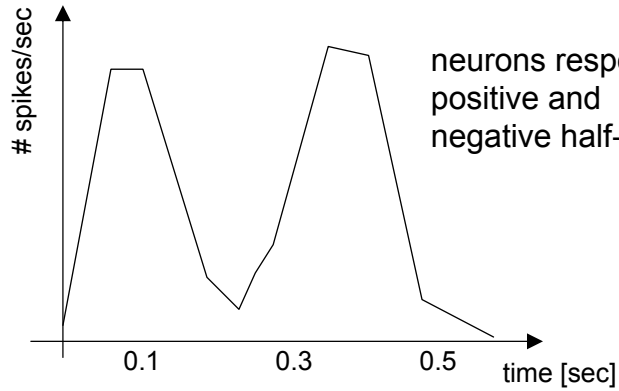
Half-wave rectification



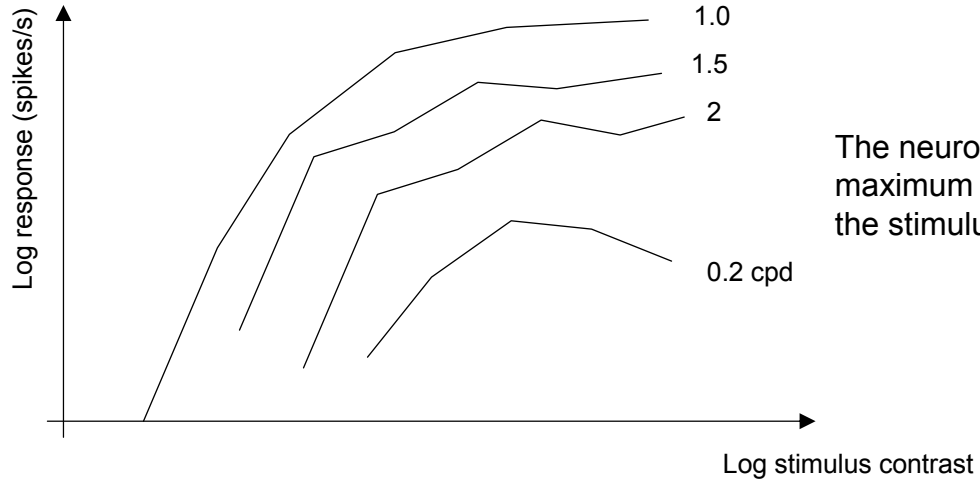
neurons do not respond to negative half-wave

Complex cells

Full-wave rectification

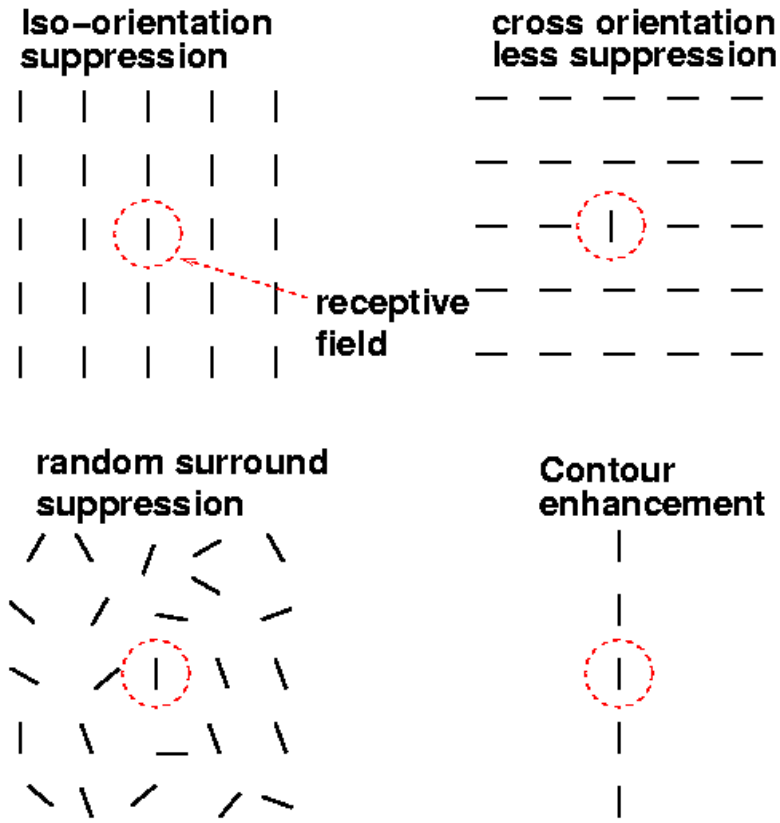


neurons respond to positive and negative half-wave



The neuron's sensitivities and maximum response depend on the stimulus spatial frequency

From local to global



[L. Zhaoping et al., *Trends in Cognitive Sciences*, Vol. 6. No. 1, Jan, 2002, p. 9-16]

Modeling receptive fields

Sparse representations and the statistics of the natural environment [olshausen]

Modeling vision

- Identification of the *optimal basis* for representing the stimuli
  - Make some assumptions on the strategies employed by the visual system
  - Derive a mathematical representation of the stimulus accordingly

- Design of the model in the feature space
  - Make some assumptions on the way the transformed coefficients are “interpreted”
  - Design a mathematical/statistical model that tries to reproduce such a behavior
  - Example: image quality metrics
- Model validation
  - Objective evaluation (*ideal observer*)
  - Subjective evaluation

Identification of the *optimal basis*

It has long been assumed that sensory neurons are adapted, through both evolutionary and developmental processes, to the statistical properties of signals to which they are exposed [Atteanave-56, Barlow-61]

Statistics of natural images ↔ Neural responses

► Efficient coding hypothesis

- "The role of early sensory neurons is to *reduce the redundancy* in the representation of the sensory input" [Barlow-61]

The simplicity of such a statement hides very difficult problems

- Description of the probability distribution over the space of natural images
  - The estimation of probability density functions on high dimensional spaces is cumbersome
- Identification of the neurons that should respond to the independent coding hypothesis
- Translation of the neural responses to *perceptual* cues

Identification of the optimal basis

- Guideline
  - Analyze the statistical properties of environmental signals (images) and derive a description of the response properties of sensory neurons based on a given *statistical optimization criterion (sparsity)*
- Method
  - Interpret the stimulus (image) as the realization of an underlying stochastic process
  - Make some assumptions on the criteria followed by the visual system for its encoding
  - Derive a cost function and an optimization rule
- Keywords
  - Sparsity & statistical independence
  - Implications of the efficient coding hypothesis

1. Statistical independence: the response of neuron N_i does not provide any information on the response of neuron N_j, for any i *different from* j
2. Sparsity: most of the neurons are not active (not responding) most of the time

Basic approximations

- Hypothesis of linearity
  - The stimulus is represented by a weighted sum of basis functions
  - Basis functions ↔ receptive fields of neurons

$$I(x,y) = \sum_i a_i \varphi_i(x,y)$$

$$\varphi_i(x,y) \quad \text{basis functions}$$

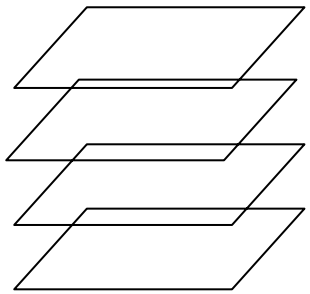
- *Stationarity*
  - The statistical parameters (mean, variance..) do not change with spatial position
- *Ergodicity*
  - Statistical properties can be inferred from a single realization (image)
- Limit to second order properties of the input statistics
  - Variance, covariance
    - Properties of points and dipoles

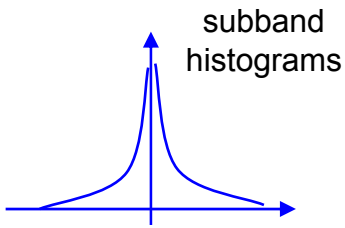
Sparse coding

- Maximization of the *sparseness* of the representation
  - The marginal histograms (i.e. the histograms of the coefficients obtained after the perceptual decomposition) have a sharp peak at zero [Olshausen&Field-96]
    - The great majority of subband coefficients are zero

Perceptual decomposition

Linear decomposition



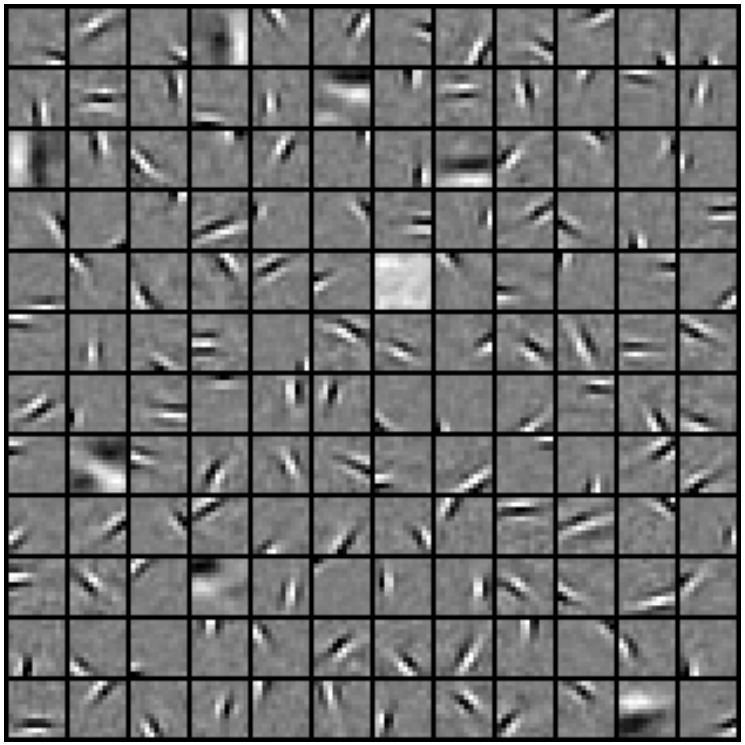


Responses of arrays of neurons which are selective for different spatial frequencies, scales and orientations ↔ *subbands*

$$I(x,y) = \sum_i a_i \varphi_i(x,y)$$

$$E(x,y) = \sum_{x,y} \left[I(x,y) - \sum_i a_i \varphi_i(x,y) \right]^2 + \varepsilon(\sigma^2)$$

Sparse coding bases



The basis functions are *multi-scale, bandpass and oriented*

Resemble RF of neurons responsible for early vision

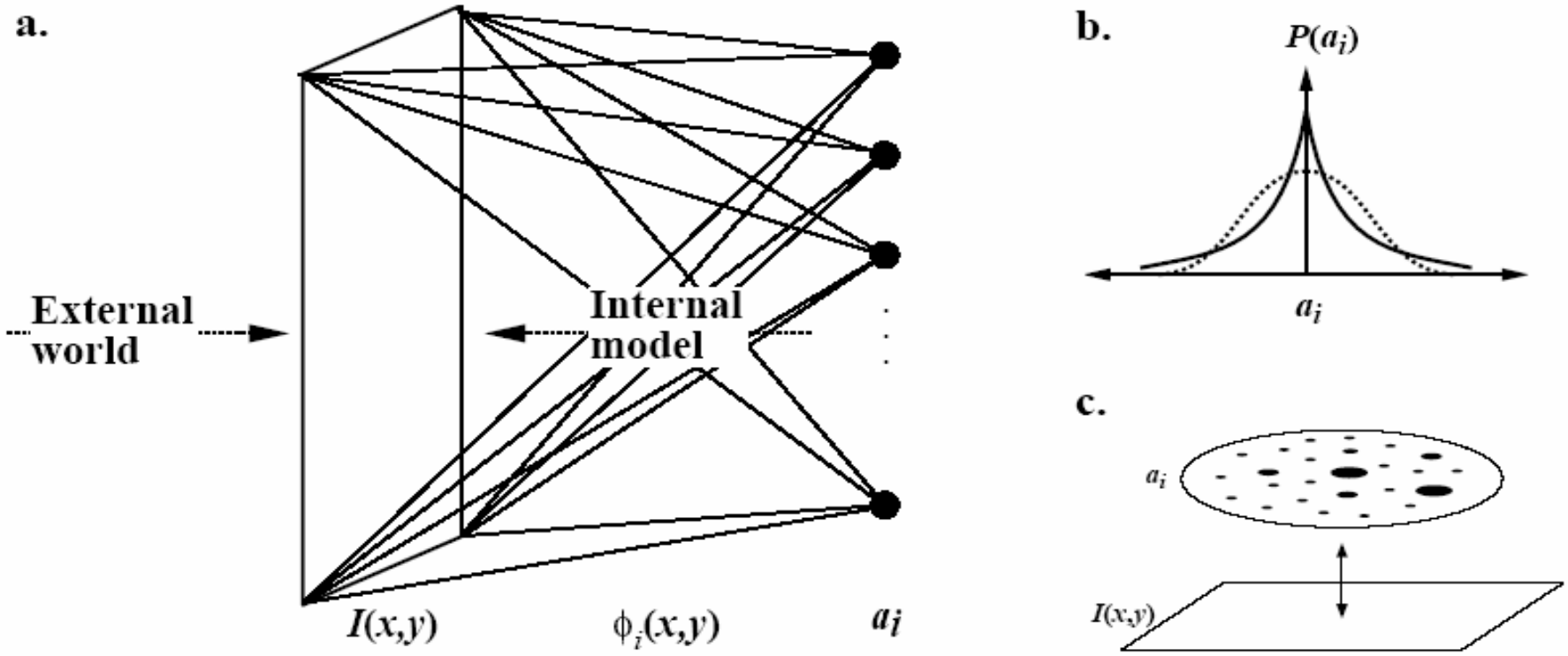
→ physiological mapping

Can be assimilated to *wavelet* families

→ SP and information-theoretic mapping

First evidence of the complementarity of the fields of vision sciences and signal processing

Olshausen's model



Olshausen's model

$$I(x,y) = \sum_i s_i \phi(x,y) + \varepsilon(x,y)$$

- For efficient coding "s_i" have to be:
  - Sparse
  - Statistically independent

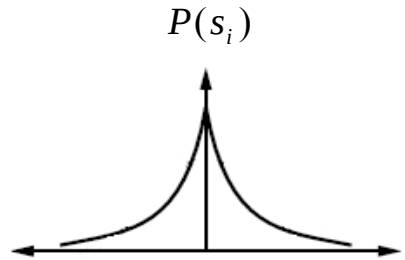
- Drawbacks of previous approaches:
  - PCA or ICA achieve the two constraints but solutions not spatially localized. Then they do not allow for overcomplete codebooks
  - Fitting Gabor wavelets functions: too many parameters to be tuned by hand

Bases-Learning Algorithm

By imposing the following probability distributions:

$$P(x \mid A, s) \propto e^{-\frac{(x - As)^2}{2\sigma^2}}$$

$$P(s_i) \propto e^{-\vartheta_i |s|}$$



it is possible to apply the Bayes' rule to derive the following cost function which trades off "representation quality" for "sparseness". Thus, the search for a sparse code can be formulated as an optimization problem minimizing the cost function:

$$E = \left| I - \sum_i s_i \phi(x, y) \right|^2 + \lambda \sum_i \vartheta_i \cdot |s_i|$$

It measures how well the code describes the image

It assesses the sparseness of the code

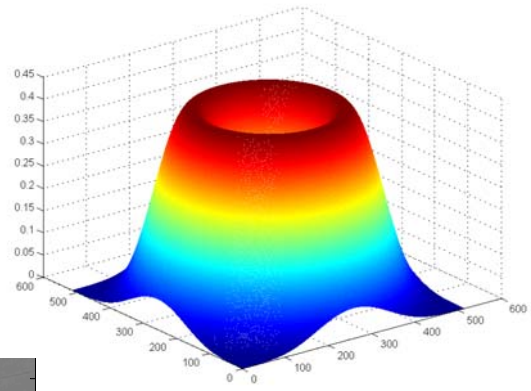
Training Sets

Each set is composed of ten images of 512x512 pixels

Preprocessing

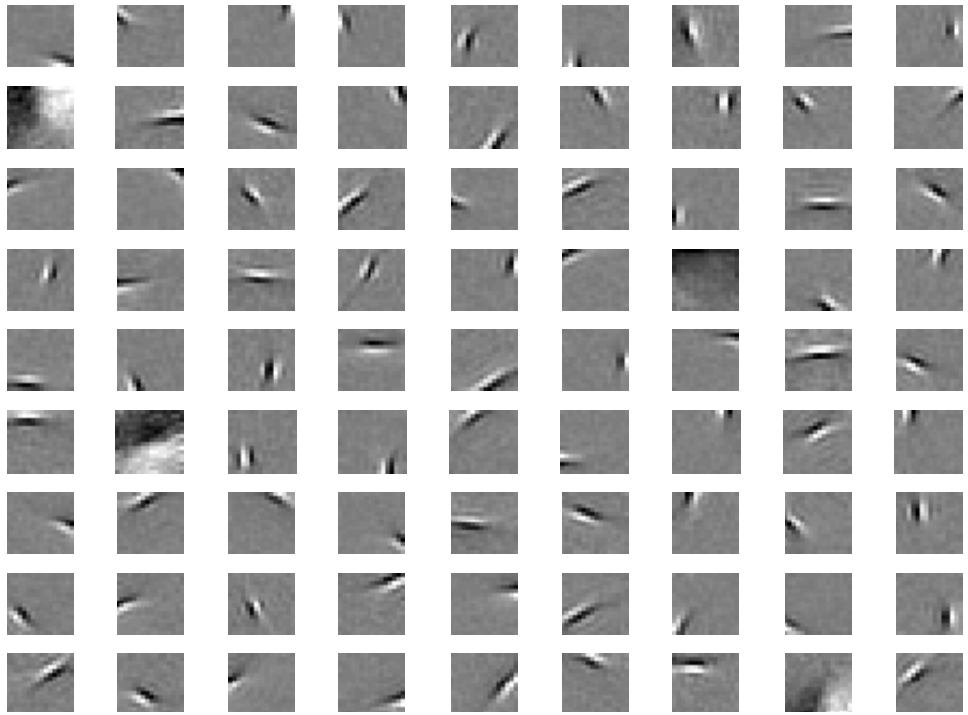
- It is needed to counteract the fact that the error computed in the cost function preferentially weights low frequencies.
- Zero-phase whitening lowpass filter:

$$R(f) = f \cdot e^{-\left(\frac{f}{200}\right)^4}$$



Result: codebook (set 1)

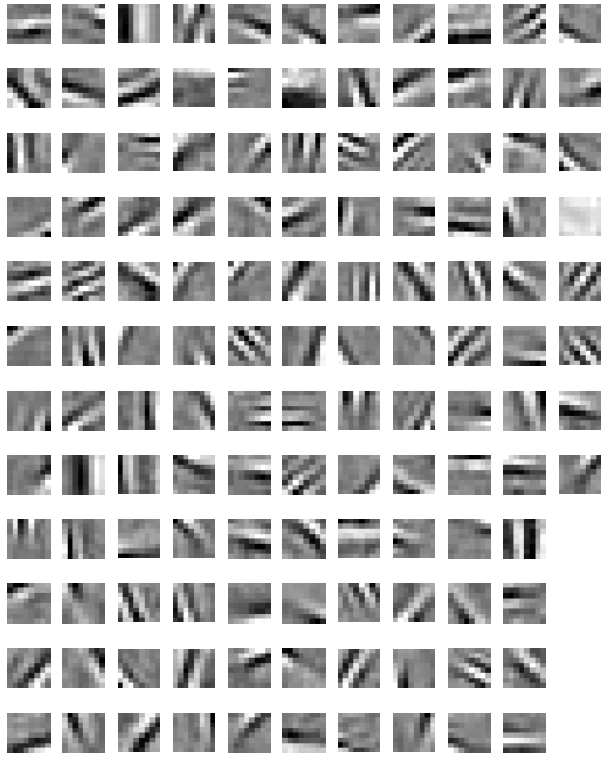
The algorithm randomly selects image patches the dimension of the chosen bases

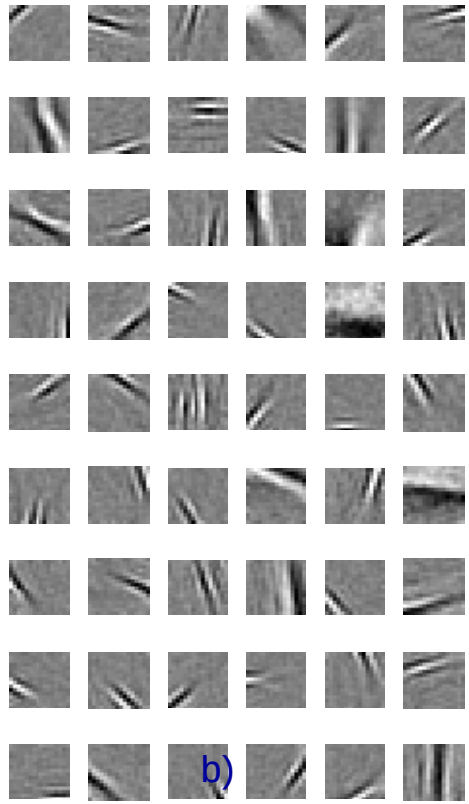


Results from training a system of 192 bases functions on 16x16 image patches extracted from scenes of nature: the results were obtained after 40,000 iteration steps (4 hours of computation)

Result: codebook (set 2)

a)

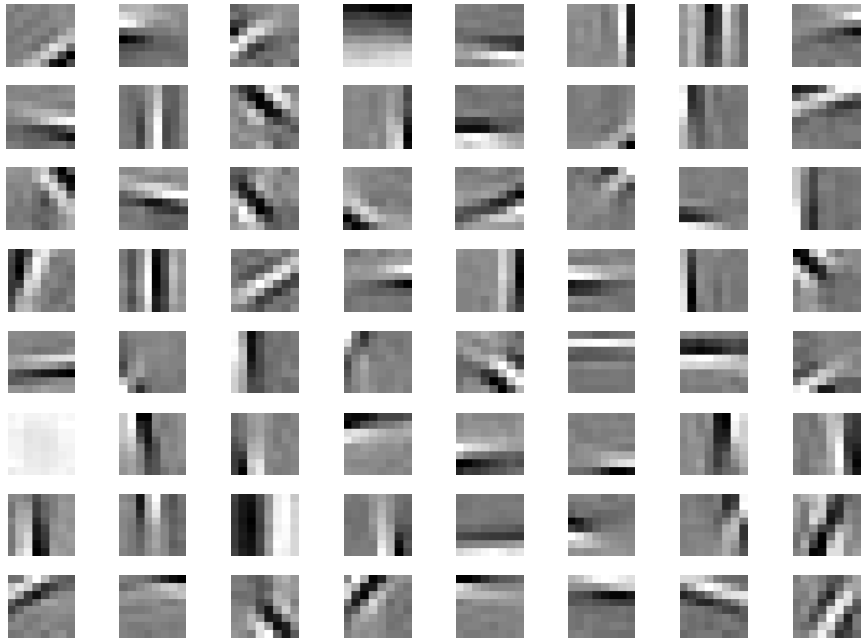




b)

a) 2x-overcomplete system of 128 bases functions of 8x8 pixels (b) 192 bases of 16x16 pixels
20,000-40,000 iteration steps: 2-4 hours of computation
The learned bases result to be oriented along specific directions and spatially well localized.
Moreover, the bases seem to capture the intrinsic structure of Van Gogh brushstrokes!

Result: codebook (set 3)

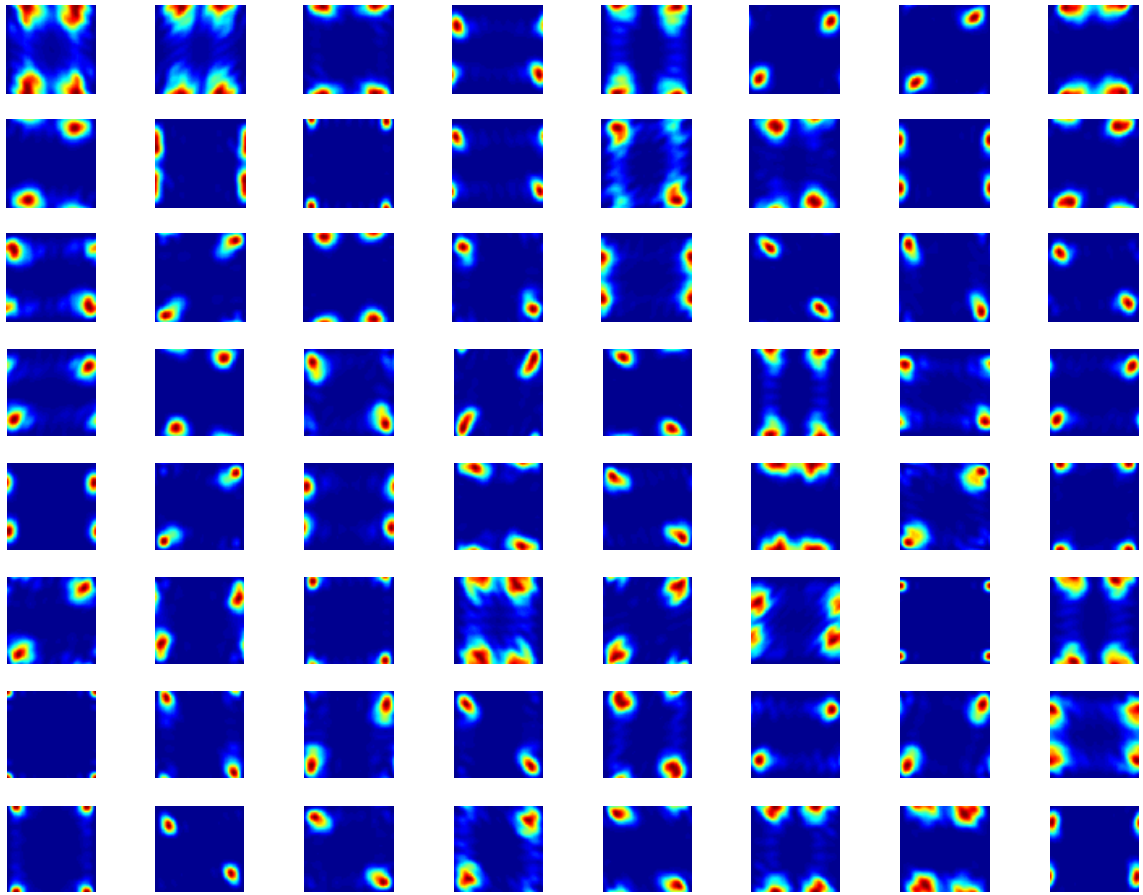


64 bases functions of 8x8 pixels

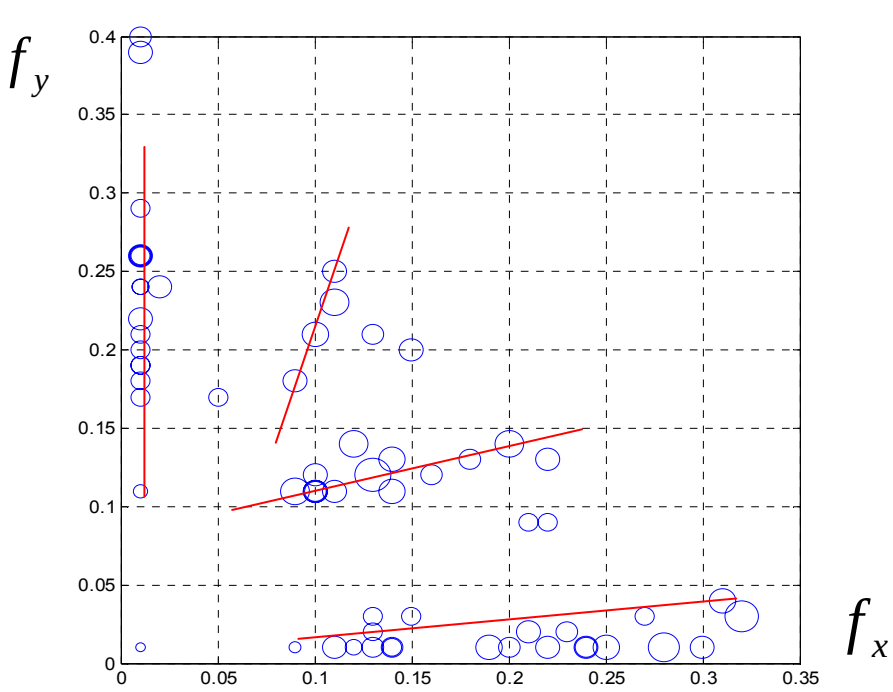
The bases seem to capture the intrinsic structure of the building elements, that are mainly composed of vertical, horizontal, slanting edges and corners.

Codebook properties

The basis functions result to be: spatially *localized, oriented* and *bandpass*



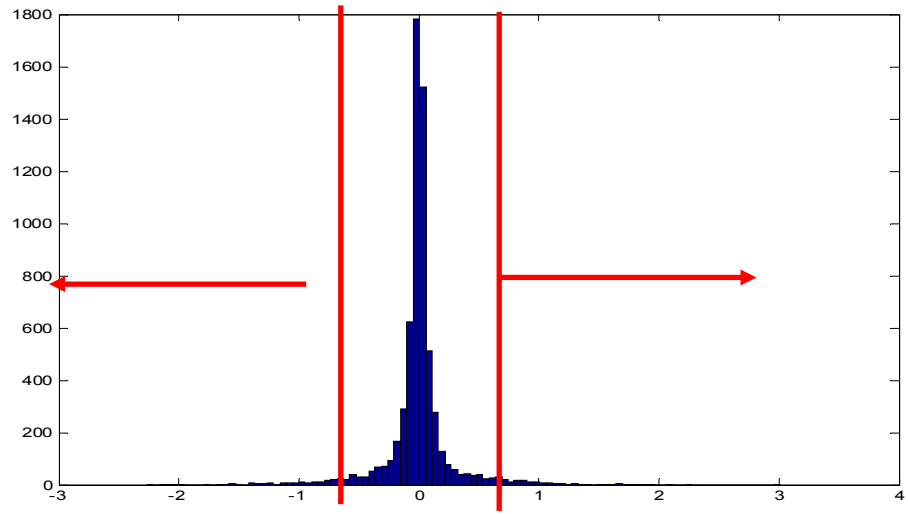
Frequency Tiling Properties



In pictures of buildings, the basis spectrums undergo certain precise directions.
These preferential directions are due to the localized orientation of the correspondent bases in the spatial domain: horizontal, vertical and slanting edges

Reconstruction

- Given the probabilistic nature of the approach, we can not have a perfect reconstruction but, conversely, the best approximation of the original picture
- At the end of the learning process, coefficient histograms undergo the Laplacian distribution imposed by the model: they are sparse!
- To have an M-bases approximation, take only the M coefficients of higher absolute value



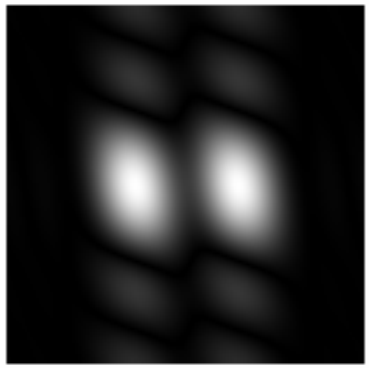
How well does the learned codebook fit the behavior of V1 receptive fields?

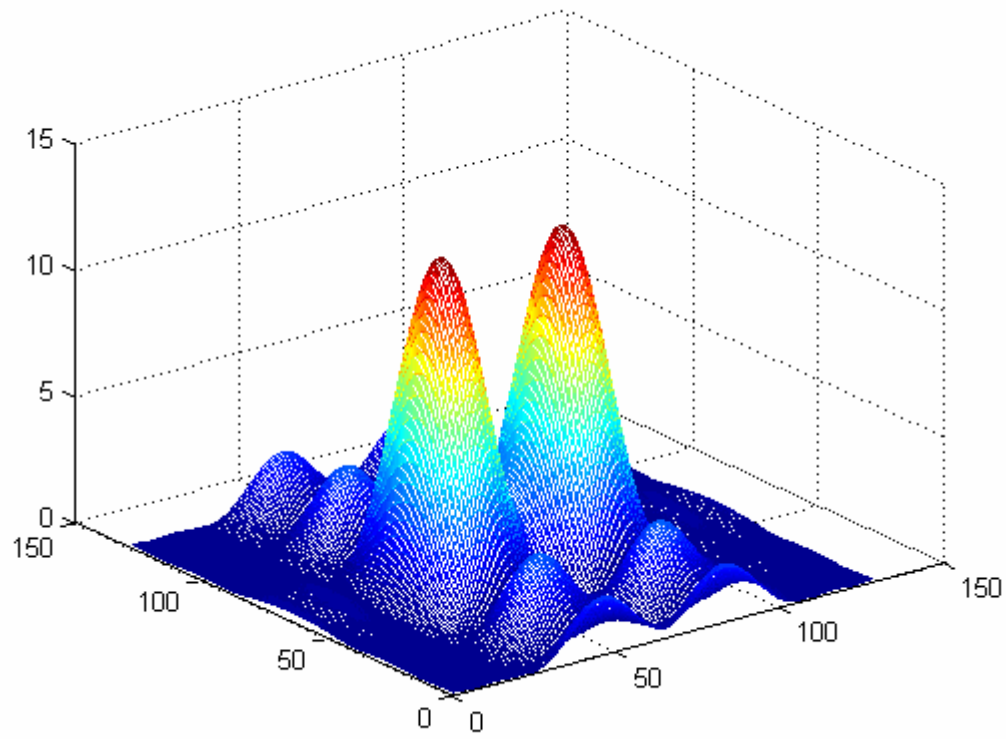
- Versus
  - Localized, oriented and bandpass bases
  - Sparseness of coefficients resemble the sparse activity of neuronal receptive fields
  - Learned bases from natural scenes reveal the intrinsic structure of the training pictures: they behave as feature detectors (edges, corners) like V1 neurons
- Against
  - Bases show higher density in tiling the frequency space only at mid-high frequencies, while the majority of recorded receptive fields appear to reside in the mid to low frequency range
  - Receptive field reveal bandwidths of 1 - 1.5 octaves, while learned bases have 1.7 – 1.8 octaves
  - Neurons are not always statistically independent of their neighbours, as it is assumed in the analytical model
- Remaining challenges for computational algorithms
  - Accounting for non-linearities as shown by neurons at later stages of visual system
  - Accounting for forms of statistical dependence

Conclusions

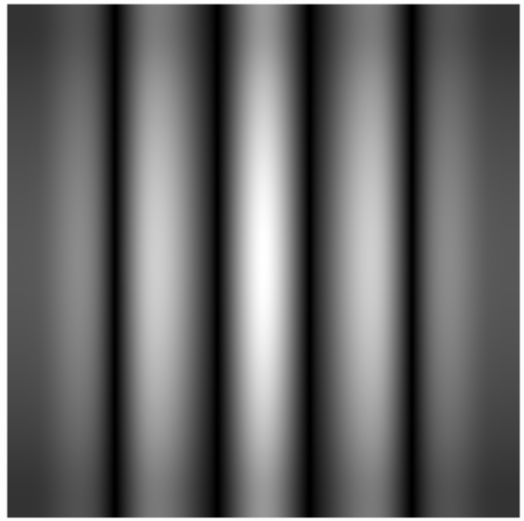
- Results demonstrate that localized, oriented, bandpass receptive fields emerge only when two objectives are placed on a linear coding of natural images:
  - That information be preserved
  - And that the representation be sparse
- The learned bases behave as feature detectors and capture the intrinsic structure of natural images
- Increasing the degree of completeness results in a higher density tiling of frequency space
- Sparseness and statistical independence among coefficients allow efficient representation of digital images
- Spatial and frequency properties of such a learned codebook reveal a lot of similarities with fitted Gabor wavelets!

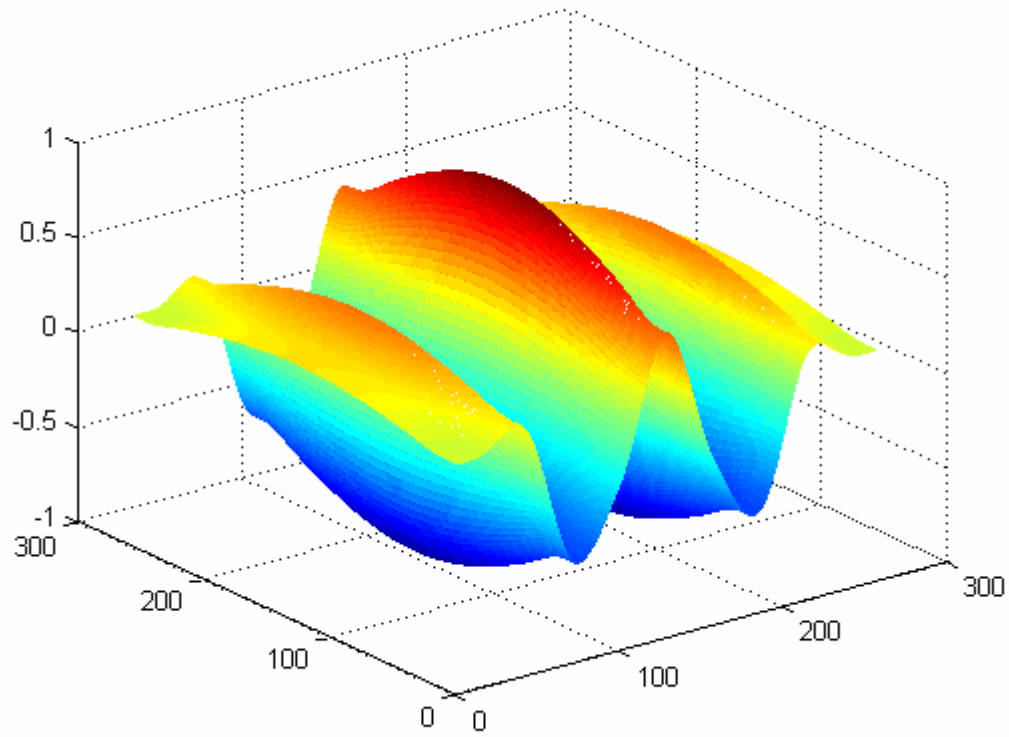
Gabor: frequency response





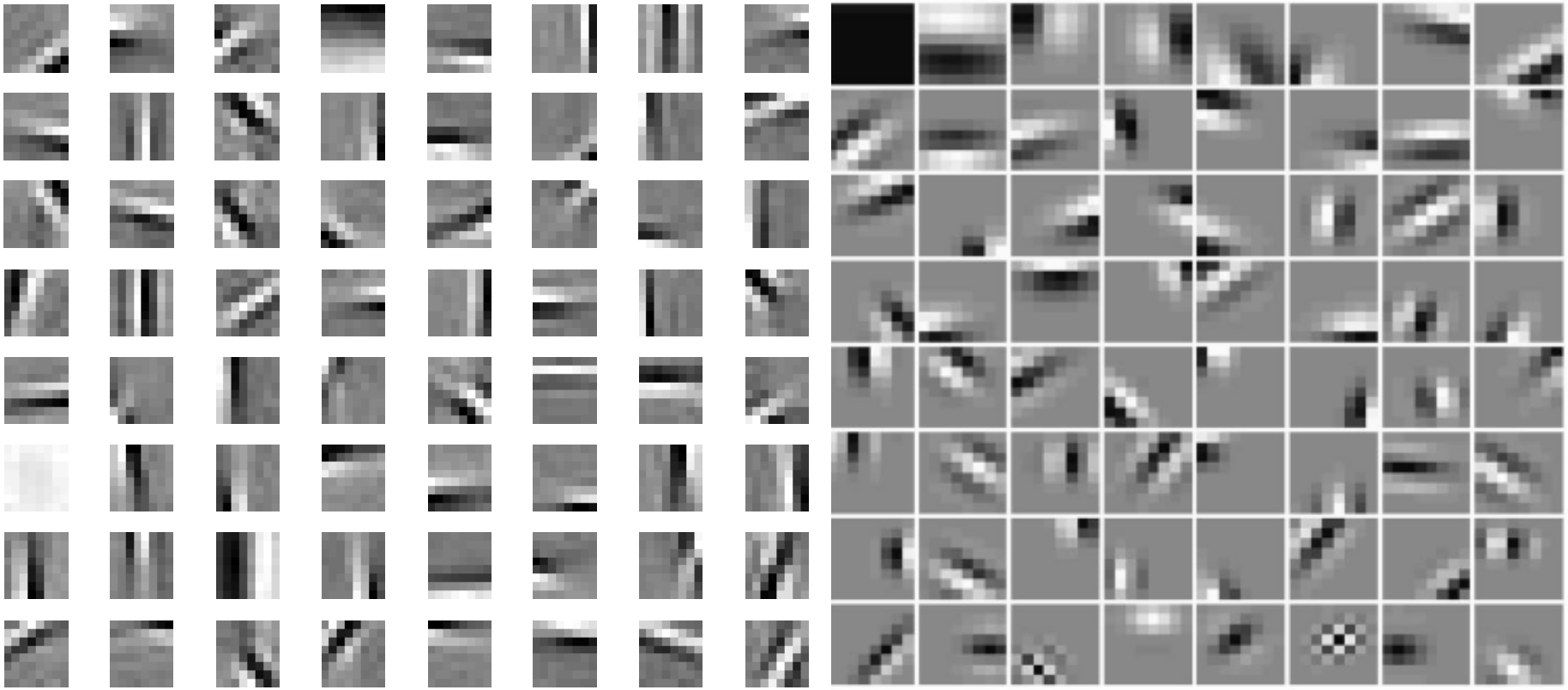
Gabor: impulse response





learned

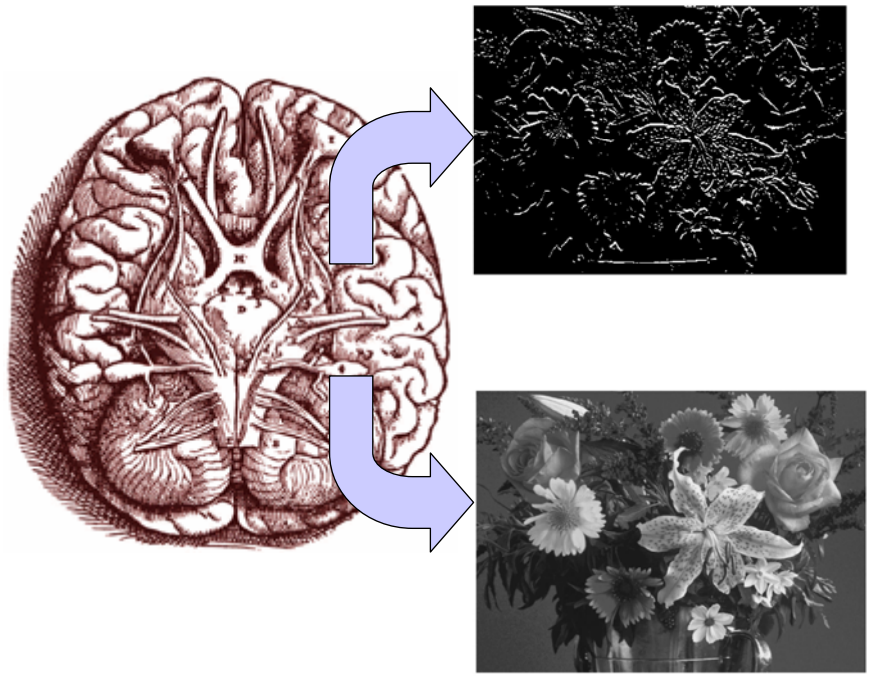
fitted Gabor



Pattern sensitivity

Pattern sensitivity

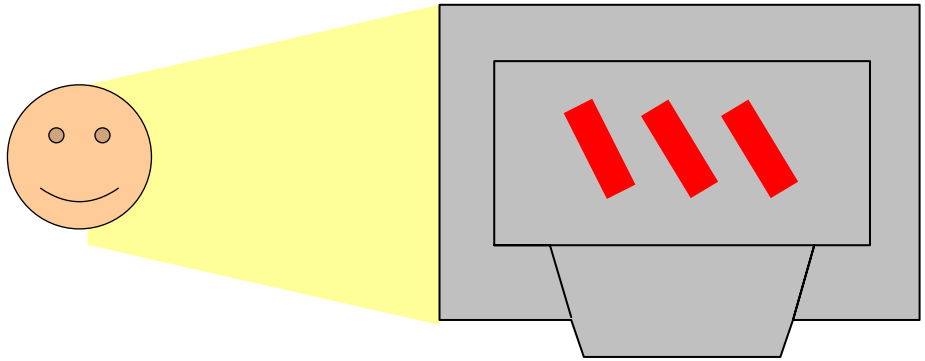
- Spatio-temporal patterns
- Behavioral experiments
  - Detection
    - Threshold
  - Discrimination
    - Just Noticeable Difference (JND)
- Neural image
  - The responses of a collections of neurons with similar RF differing in the spatial position make up a neural image
    - Each neural image is representative of a *population* of neurons which encodes a different property of the stimulus.
    - The **intensity** of the neural image at a point represents the **activity level** of the corresponding neuron

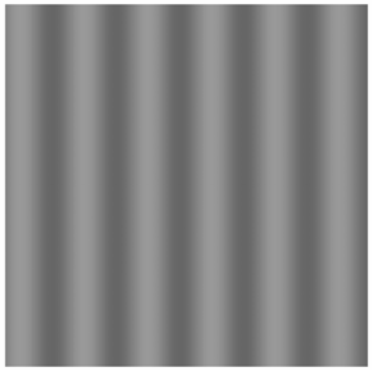


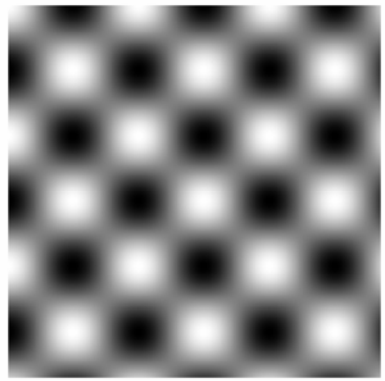
Pattern sensitivity

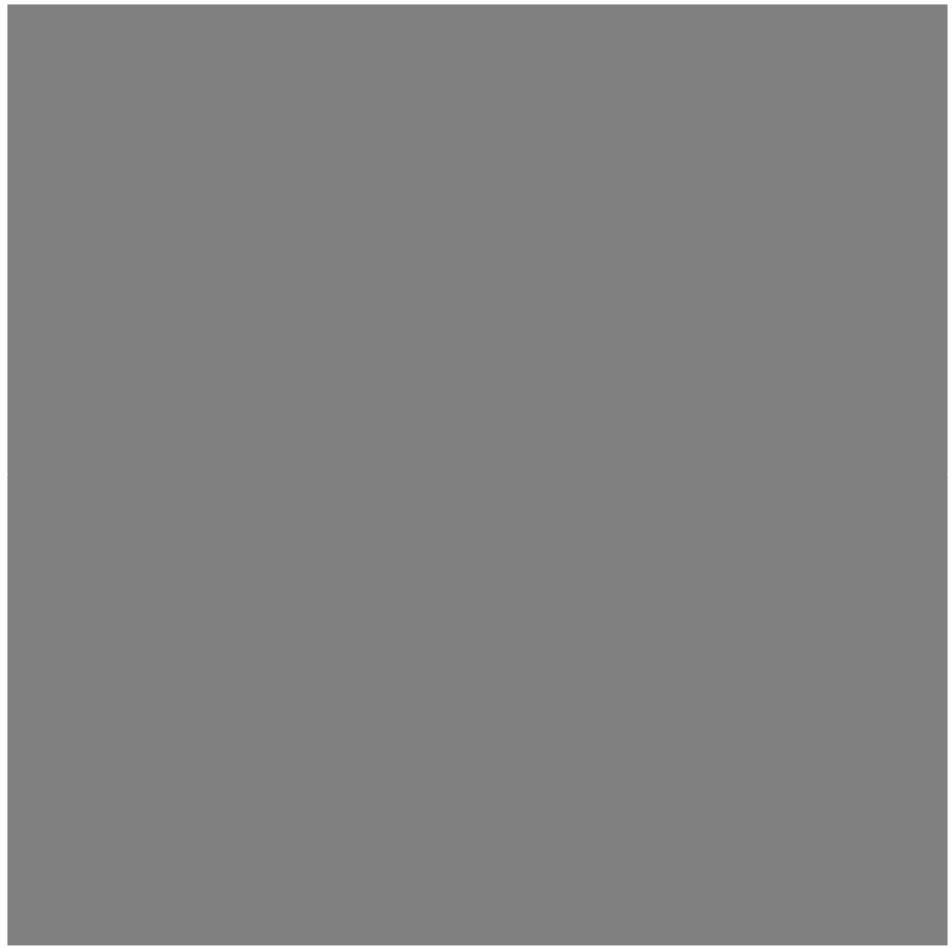
- Psychophysical investigation
  - Threshold experiments
  - Discrimination experiments
- Generalization of the definition of the spatiotemporal CSF as resulting from the whole vision processes a stimulus is subject to in the psychophysical investigation loop
  - No longer a way to characterize the responses of neurons (either completely for those for which linearity holds – LGN neurons and simple cells of V1 – or partially – complex cells in V1) BUT one of the many measures aiming at characterizing the visual system
- Pattern Contrast Sensitivity
- Pattern Masking
- Towards a multiresolution representation

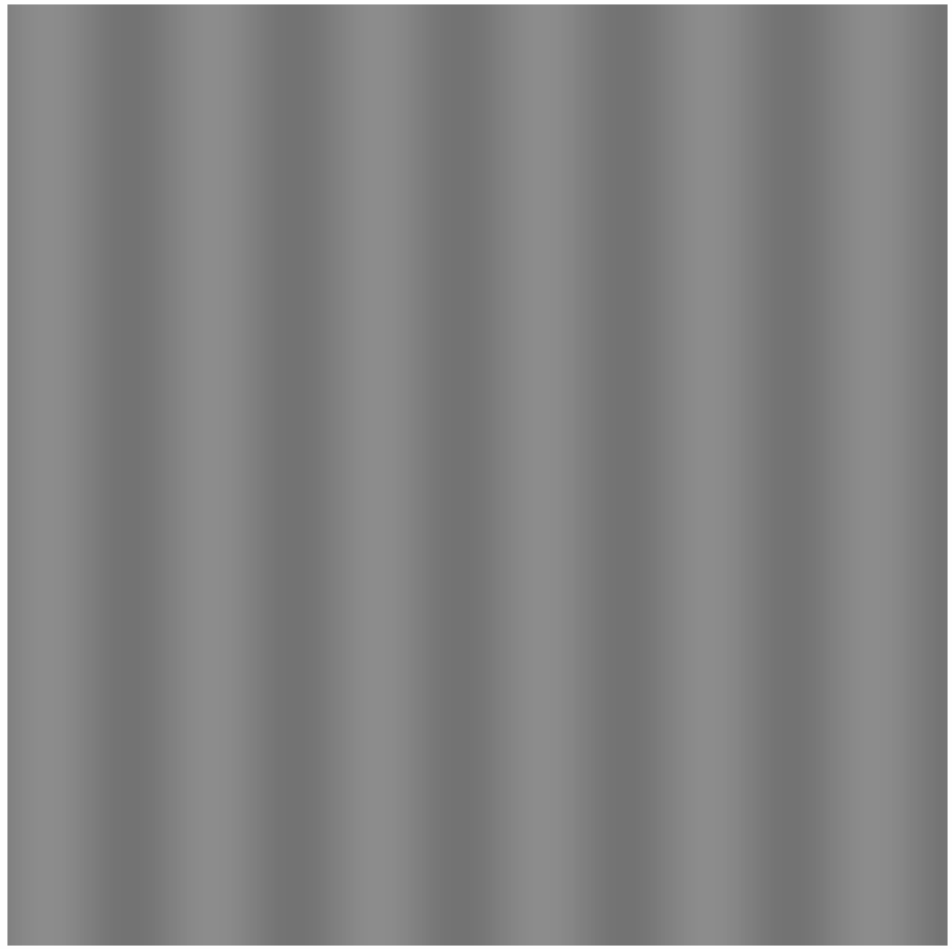
Spatial CSF

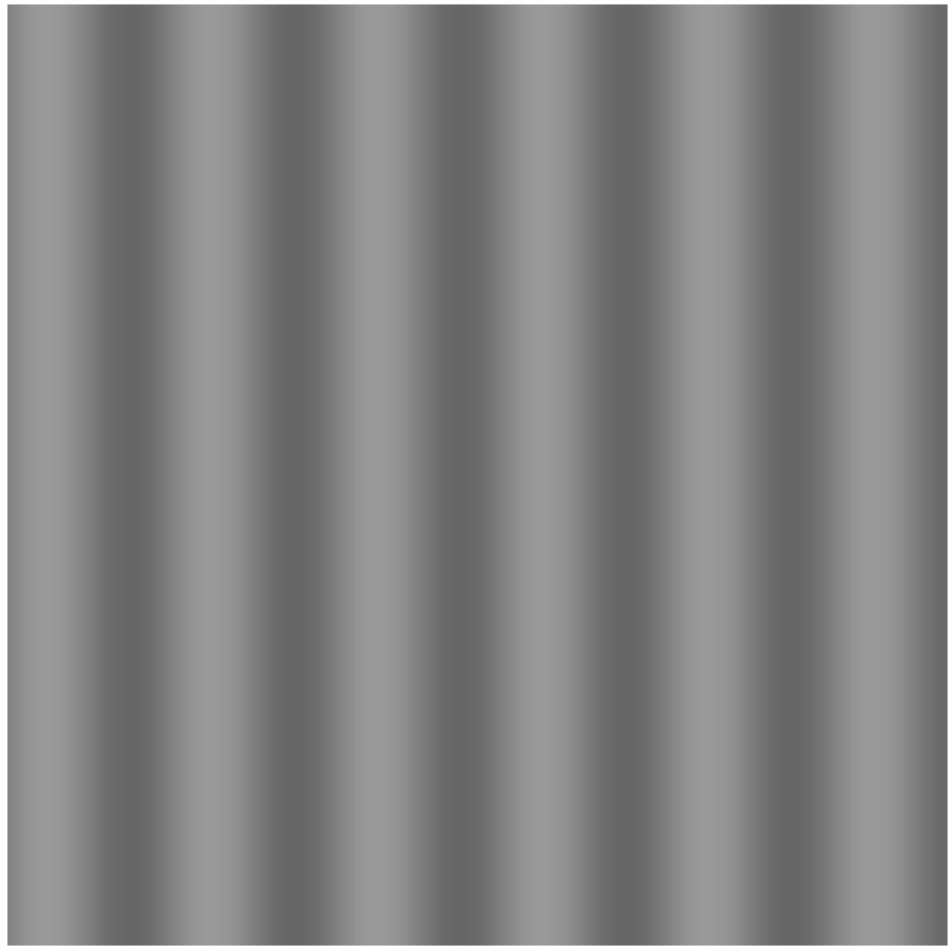




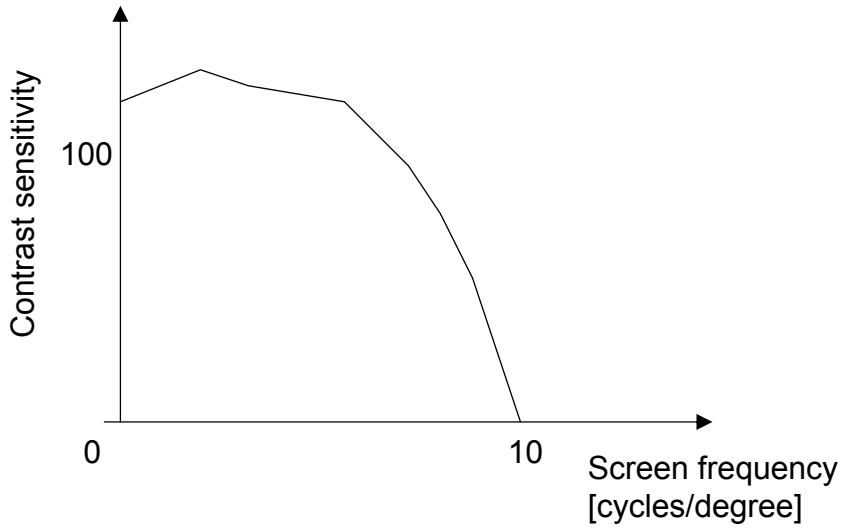








Spatial CSF



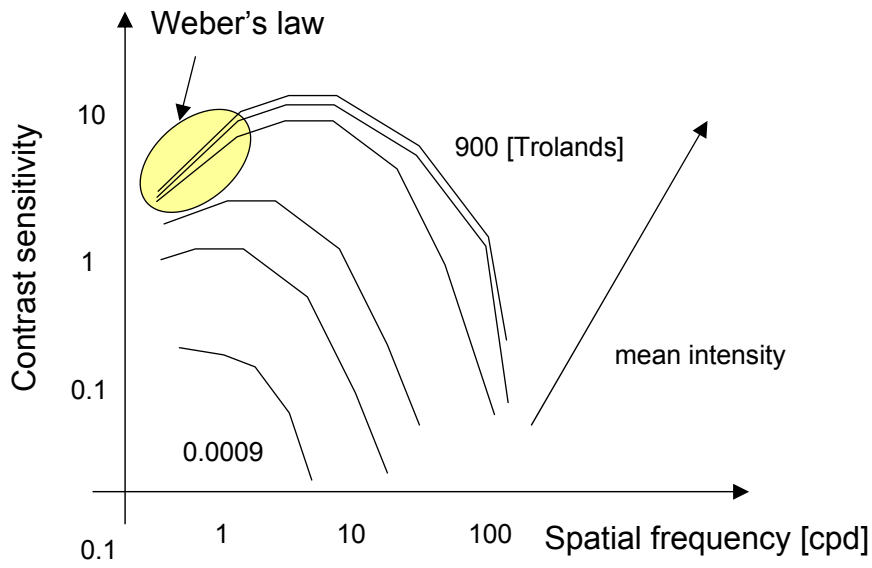
- Fixed experimental conditions
  - Given background intensity level
  - Monochromatic stimuli (sine waves)
- Parameters
  - Stimulus contrast
  - Spatial frequency
- Measure
  - Contrast at threshold c
  - or Contrast sensitivity 1/c
- Threshold estimation
  - For each tested frequency, the value of the contrast for which the stimulus becomes visible are collected
  - The contrast at threshold is defined as the smallest one

Features

• Low sensitivity at high spatial frequencies
• No improvements at low frequencies; small fall near zero, probably due to neural factors
• *Band-pass* structure

• Good agreement with the neural CSFs

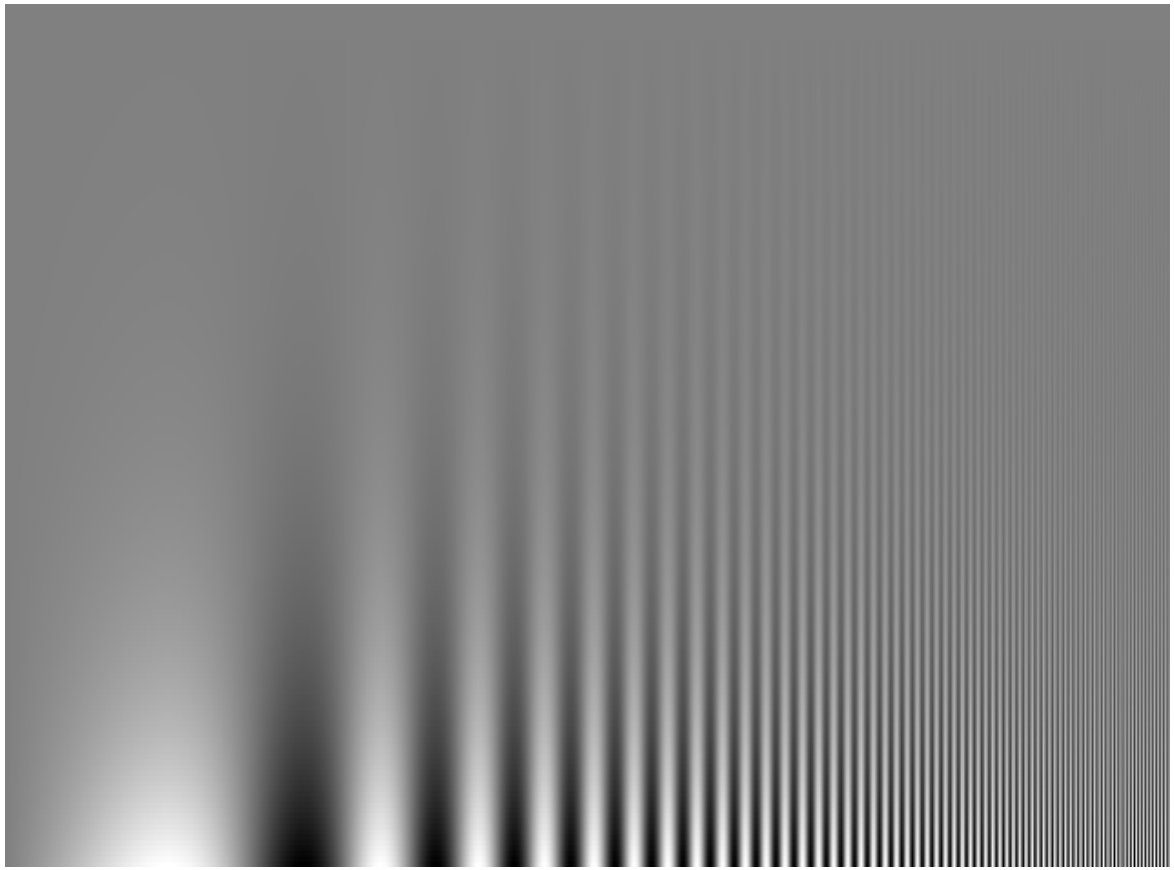
Light adaptation



Stimulus: monochromatic light at 525nm

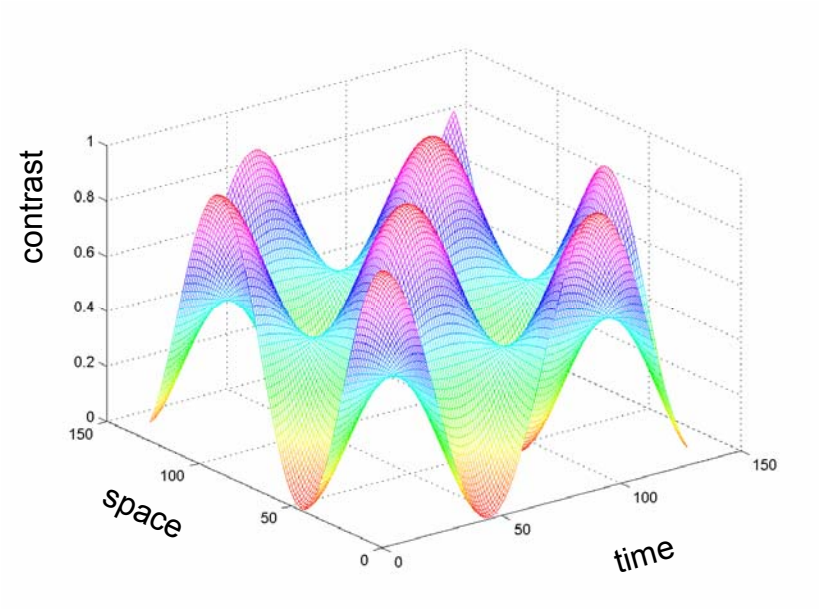
- Low light levels
  - Integration across many neurons to achieve a reliable signal → compromises spatial resolution
- High intensity levels
  - No need to integrate → improved spatial resolution
- Webers' law regime
  - Range in which the contrast sensitivity becomes constant
    - Low spatial frequencies
  - Kernel of truth: contrast sensitivity varies of 2 orders of magnitude versus the 6 of the mean background illumination

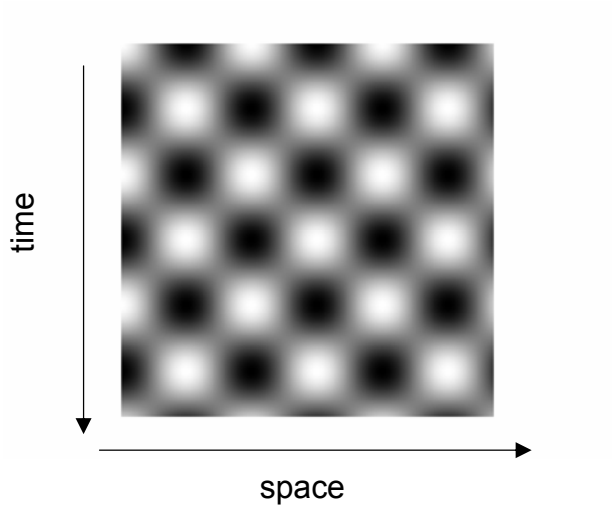
Find your CSF!



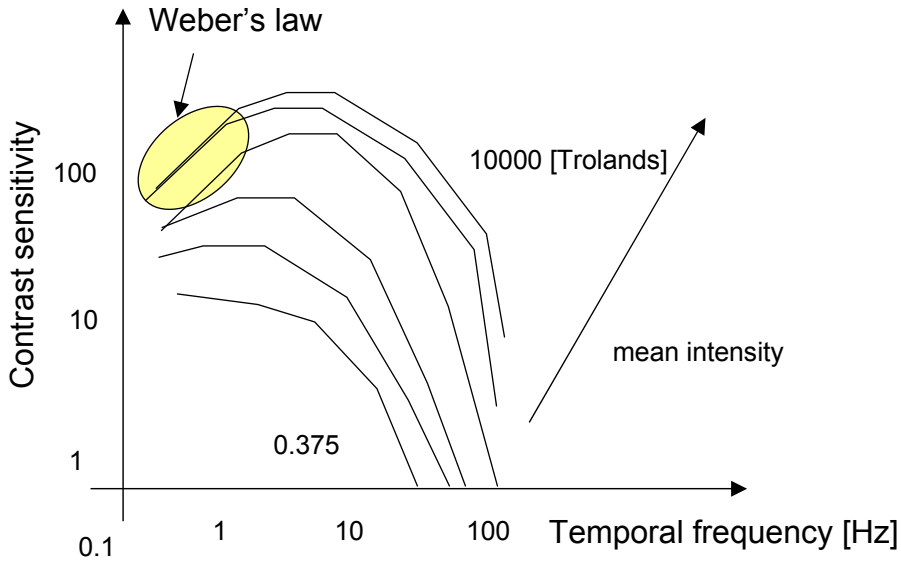
Spatio-temporal CS

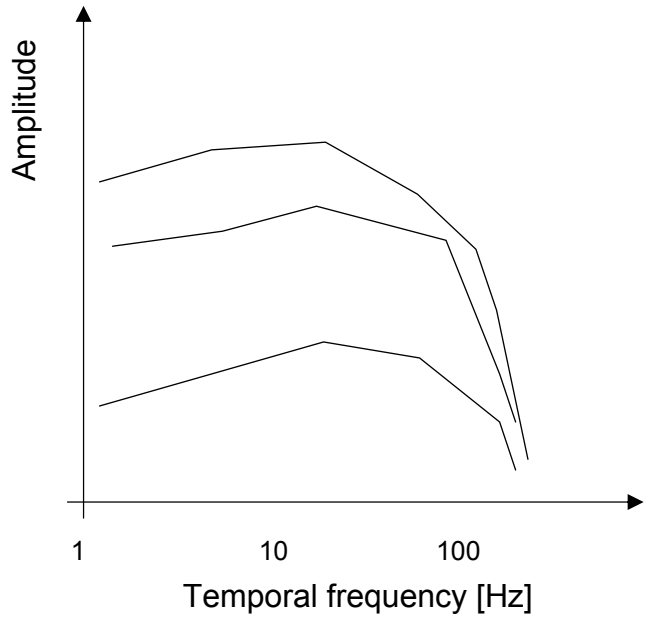
- Flickering stimuli
  - The temporal frequency of the flicker f_t is an additional control variable
  - For a single neuron, the responses to many repetitions of the stimulus must be collected and averaged → *peri-stimulus time histogram* (PSTH)
- Space-time receptive fields





Temporal sensitivity

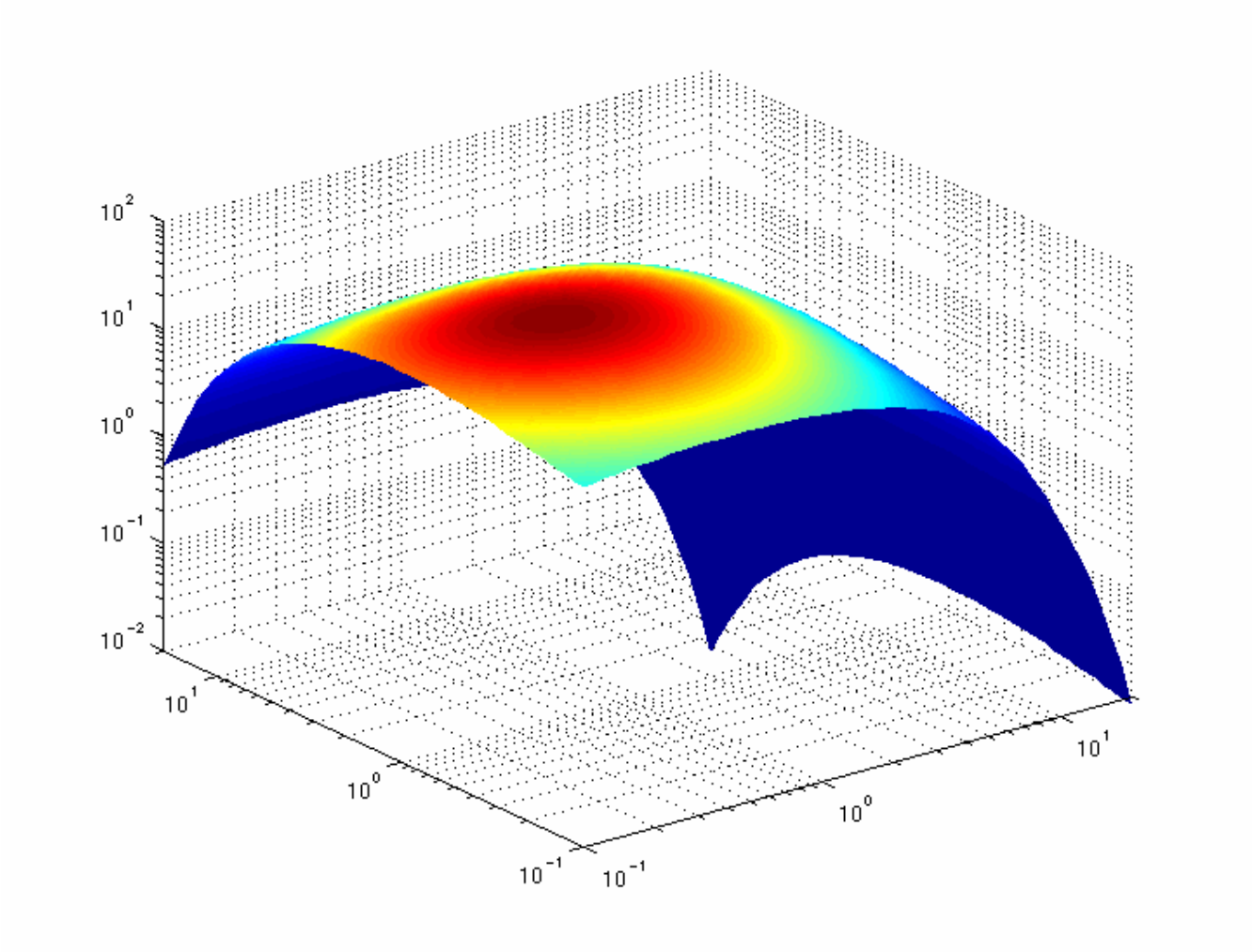




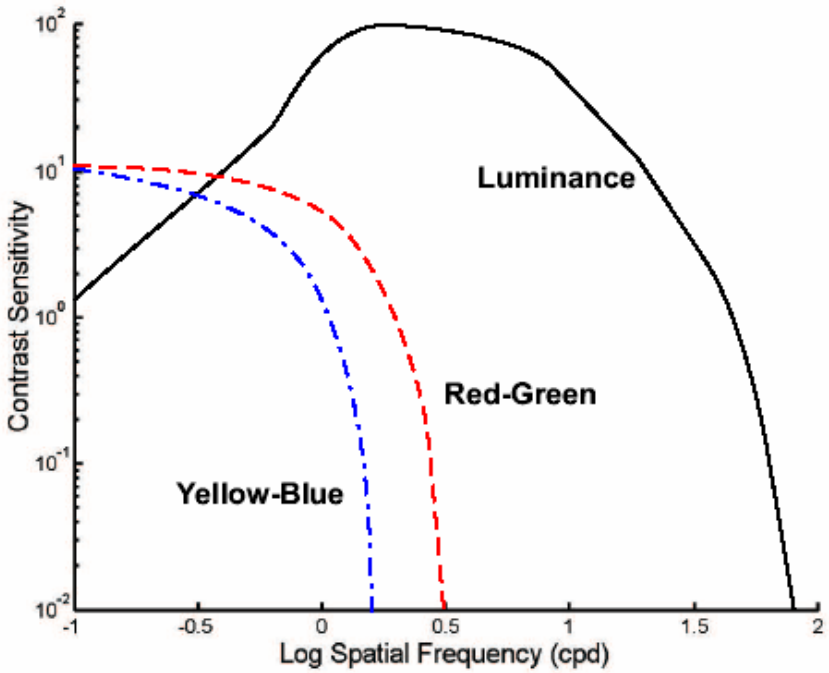
Weber's law holds above about 5 Trolands

2 degrees of visual angle

Spatio-temporal CSF



Chromatic CS



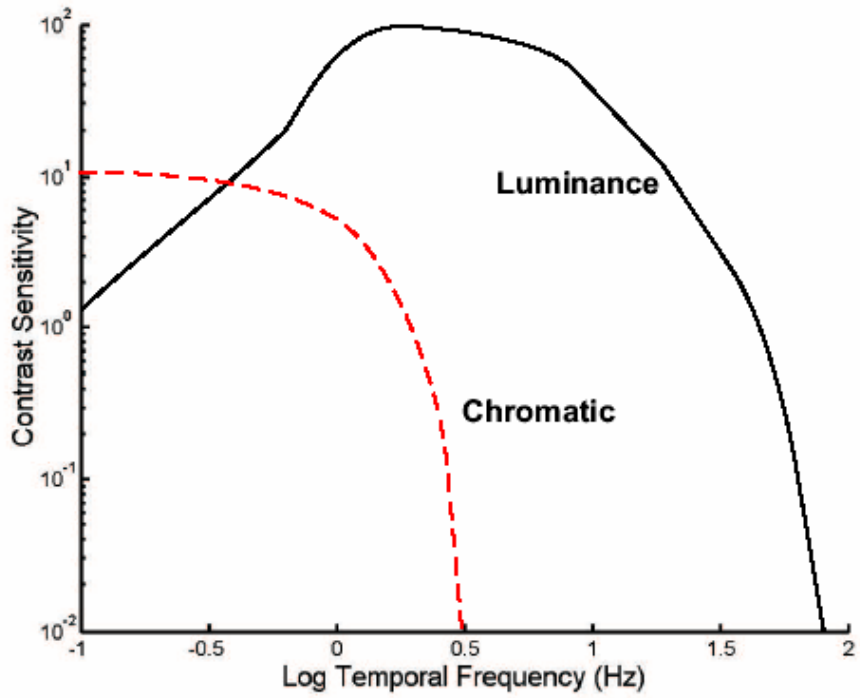
The spatial CSF for luminance contrast is band-pass in nature, with a peak sensitivity around 5 cycles per degree. This function approaches zero at zero cycles per degree, illustrating the tendency for the visual system to be insensitive to uniform fields. It also approaches zero at about 60 cycles per degree, the point at which details can no longer be resolved by the human eye.

The chromatic mechanisms are of a low-pass nature and have *signicantly lower* cut-o frequencies. This indicates the reduced availability of chromatic information for the details.

The yellow-blue CSF has a lower cutoff frequency then the red-green one, because of the pattern of S cones in the retina.

Note also that the luminance CSF is much higher than the chromatic CSFs. This denotes a greater sensitivity of the visual system to small changes in luminance contrast compared to chromatic contrast.

Chromatic temporal sensitivity

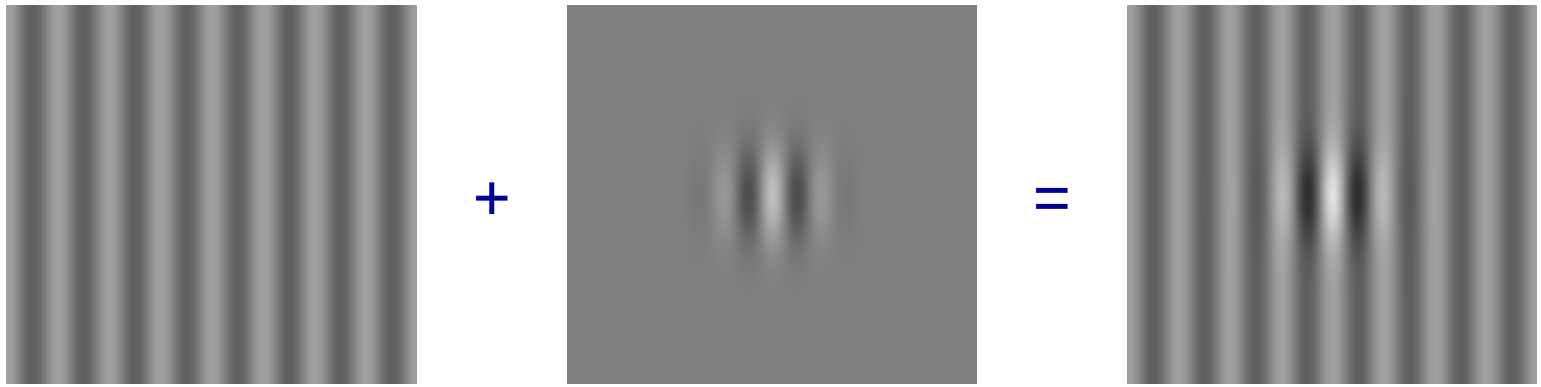


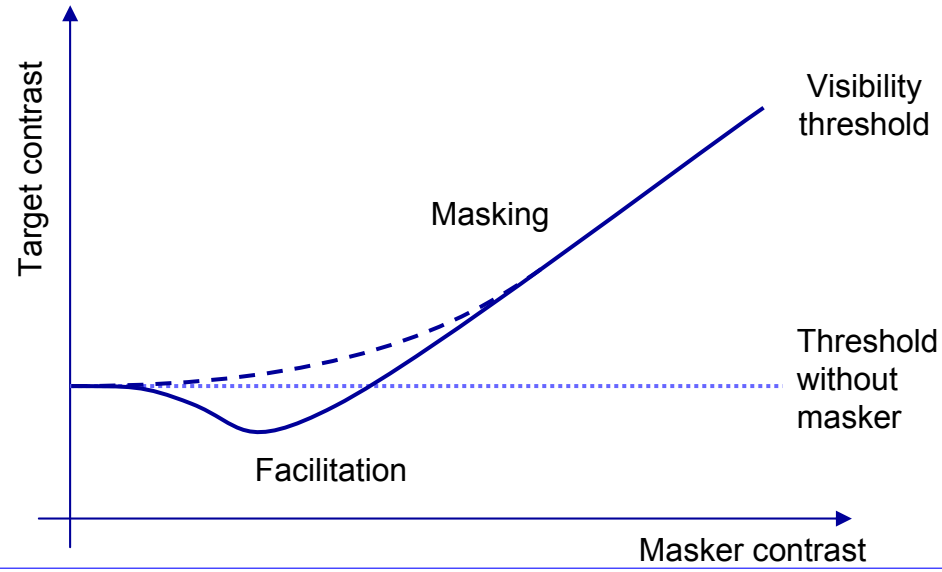
Typical temporal CSFs for luminance and chromatic contrast. They share many characteristics with the spatial CSF.

Luminance temporal CSF is still higher in both sensitivity and cut-off frequency than are the chromatic temporal CSFs.

It also exhibits band-pass characteristics, while chromatic temporal CSFs have low-pass behavior.

Pattern masking





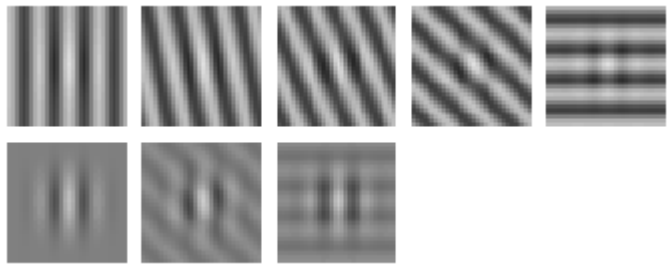
Pattern masking

Some hints

- The presence of the masker facilitates detection at low masker contrast and masks detection at high masker contrast
  - *Dipper effect*, typical of measurements at threshold, like signal visibility in noise
- When the frequencies differ by a factor of 2 there is still masking but not significant facilitation
- When the frequencies differ by a factor 3, the effect of masking is reduced

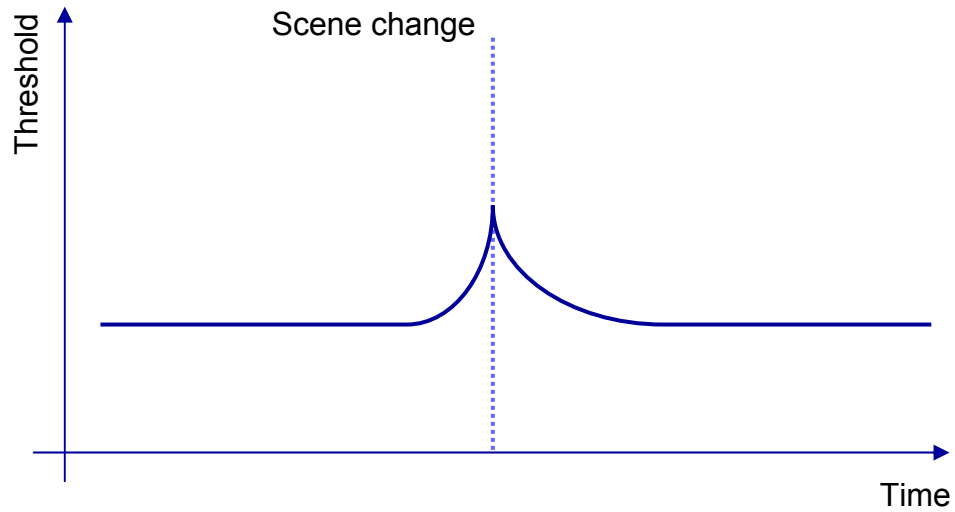
Main limitations

- Only simple stimuli
  - Monochromatic signals
  - One signal and one masker
- No models available to predict the masking efficiency of non monochromatic maskers
  - Not even maskers consisting of the sum of 2 sinewaves
- Luminance domain



Temporal masking

- Presenting one visual stimulus (a "mask" or "masking stimulus") immediately after another brief (≤ 50 ms) "target" visual stimulus leads to a failure to consciously perceive the first stimulus. A similar phenomenon can occur when a masking stimulus precedes a target stimulus rather than following it: this is known as forward masking
- Masking behavior depends on
  - Stimulus type (grating/noise)
  - Orientation, frequency, color, …
- Temporal masking
  - Sensitivity drop around scene changes



Why we are interested in vision?

Issues

- Automatization of all the tasks based on vision
  - Image segmentation, classification, pattern recognition
- Automatic assessment of image quality
- Compression and coding
- Security and watermarking
- Design of new imaging devices
  - From capture to display

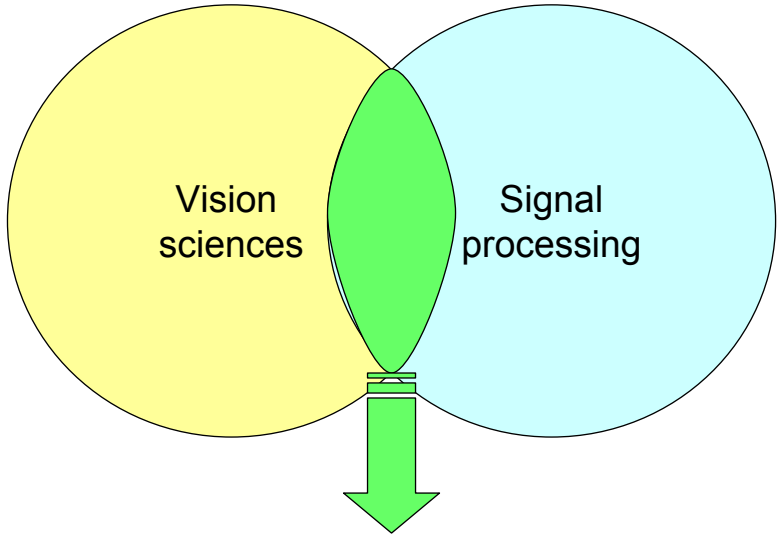
Hints&guidelines

- Exploitation of the CSF
  - Perceptual quantization
- Exploitation of the masking properties
  - Compression (as before)
  - Watermarking
- Exploitation of the color perception mechanisms
  - Compression
  - Watermarking
  - Design of capture devices
  - Image rendering

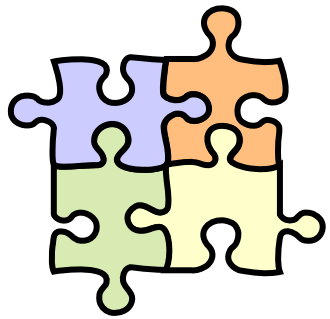
On top of this: we need a framework for the *effective representation* of the visual stimulus

Core issue

- Can we learn something about this from the investigation of visual processes?



Multiresolution representations



References

- *Foundations of Vision*, B. Wandell, Sinauer Associates Inc. Publishers, Sunderland Massachusetts
- *The first steps in seeing*, R.W. Rodieck, Sinauer Associates Inc. Publishers, Sunderland Massachusetts
- *Color Science, Concepts and Methods, Quantitative data and Formulae*, Wyszecki and Stiles, Jhon Wiley and Sons Inc.
- *Color Vision and Colorimetry, Theory and Applications*, D. Malacara, SPIE Press