# 01 LINEAR SYSTEMS

If an input-output system is linear we can use linear systems methods to characterize it. If successful, a small number of measurements allows us to predict the response to any input.
To test whether the system is linear, we need to test its response, *r*, to a number of inputs:

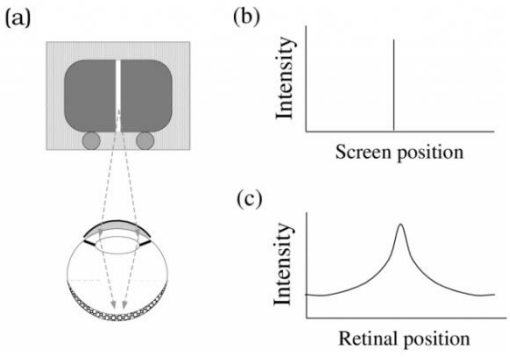
1. homogeneity (= proportionality): $r(\alpha x) = \alpha r(x)$

2. additivity: $r(x_1 + x_2) = r(x_1) + r(x_2)$

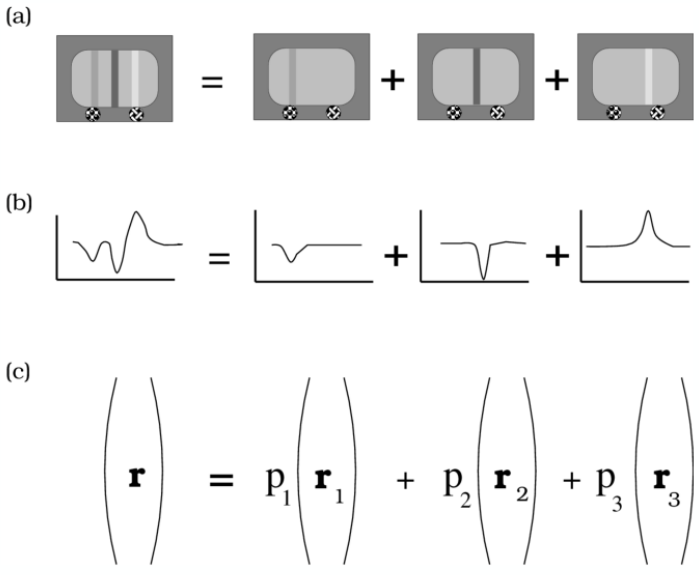
Superposition is the fulfilment of both homogeneity and additivity: $\alpha r(x_1) + \beta r(x_2) = r(\alpha x_1 + \beta x_2)$

- Any derivative, or combination of derivatives of any order
- An integral expression.
- A convolution with some fixed waveform.
- Any combination or concatenation of the above.

## Applied to Optics:




## Homogeneity (proportionality) and Addivity



## Shift Invariant

If the response r to an input x is the same over time (no memory!) and/or space, then the system is shift-invariant → if we measure:

r(x) at one position, and now shift the input x (in space and/or time), we get the same r but shifted (in space and/or time)

In our previous example: measurement of r1 would have been enough, we can deduce the individual responses r2 and r3 from r1 (shift-invariance & homogeneity) as well as the total response (additivity).

Sinewaves are the eigenfunctions of a shift-invariant linear system. Which of the following mathematical statements express this fact correctly?

- ◉ $f[\alpha \sin(2\pi\omega t + \psi)] = \beta \sin(2\pi\omega t + \phi)$
- ○ $f[\alpha \sin(2\pi\omega t + \psi)] = \alpha \sin(2\pi\omega t + \phi)$
- ○ $f[\alpha \sin(2\pi\omega t + \psi)] = \sin(2\pi\alpha\omega t + \phi)$
- ○ $f[\alpha \sin(2\pi\omega t + \psi)] = \beta \sin(2\pi\omega t + \psi)$
- ○ $f[\alpha \sin(2\pi\omega t + \psi)] = \alpha \sin(2\pi\omega t + \psi)$
- ○ $f[\alpha \sin(2\pi\omega t + \psi)] = \beta \sin(2\pi\alpha\omega t + \phi)$

## Eigenfunctions

Consider the linear system L. Every linear system has **a set of special inputs $x_0$** such that **the response r of the system to the stimulus $x_0$ is simply a (scaled version) of the input**:

**$r = L(x_0) = \lambda * x_0$**

- **shift-invariant linear systems (SILS)**
  - Eigenfunctions are **Sine waves** (sin and cos)
- when analysing SILS, all that changes is the **amplitude and phase of the sine wave** going into a SILS

## Applied to Optics:

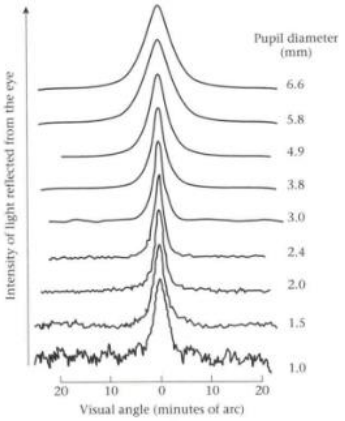
Visual system (approx.) SILS therefore such things like **Double Passage** can be calculated out

raw data:

1. small pupil diameter means less light on the retina: noisy measurements

2. large pupil diameters show more blur — lens imperfections (aberrations) show up

3. double-pass: the light measured had to go through the optics twice

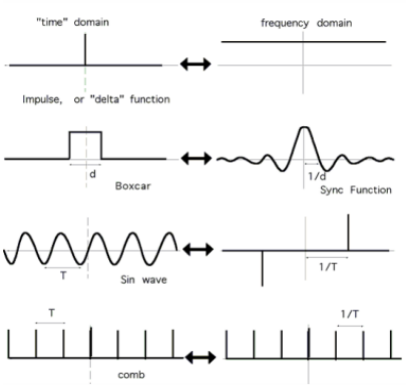



## Campbell & Gubisch (1966): Human Linespread Function

estimated LSF of human eye using double pass method

- for small pupils the retinal image quality was limited by **diffraction**

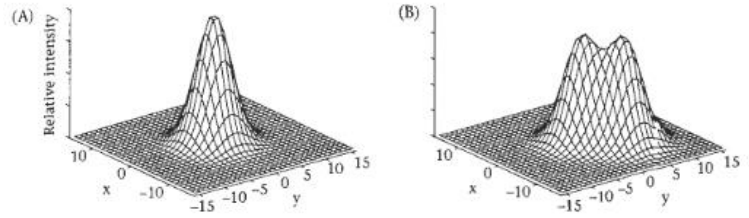
## Fourier Transform

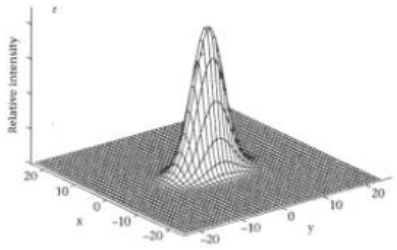



The Fourier transform is a mathematical transform that expresses a mathematical function of time/space as a function of frequency.

**Modulation Transfer Function (MTF):** the magnitude response of the optical system to sinusoids of different spatial frequencies, also the FT of PSF / LSF

## Pointspread and Astigmatism




PSF, LSF describe spread(blurring) induced by an optical system on a point or line

from PSF --> LSF
from LSF --> PSF if assumed symmetric

**Astigmatism**:
An optical system that demonstrates astigmatism is **one whose point spread function is asymmetric** . This means that the system's linespread function will depend on the line orientation .

## Depth of Field & Accomodation

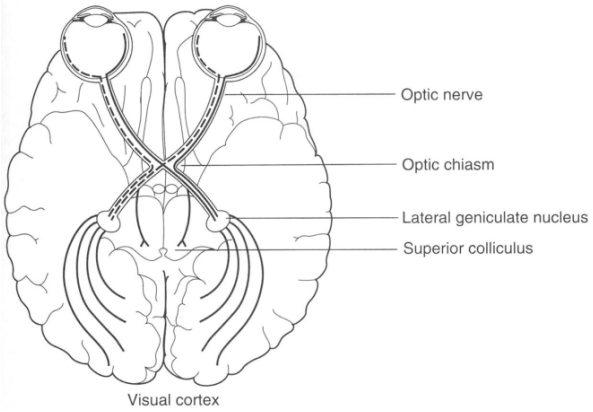
The optical power of a lens is a measure of how strongly a lens bends the incoming light (refraction) Measured in diopters, which is the reciprocal of focal length in meters. The average human eye focusing a source at optical infinity (6m) onto the retina has a distance from mid-cornea to retina of 0.017 m. 1/0.017 = 58.8 diopters.

## Chromatic Aberration of human eye

Final source of image degradation --> chromatic aberration

Cornea and lens refract the light to focus on the retina. Refraction, however, is a function of the wavelength of light. Only one wavelength can be in focus at a time.

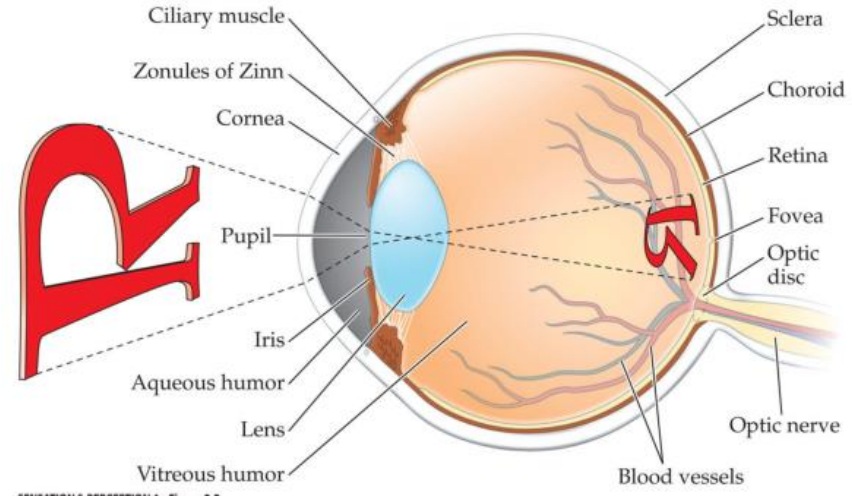
# 02 Human Eye & Retina




Frontal, Parietal, Temporal, Occipital Lobe, Extrastriate, Striate Cortex

Light -> Eye -> optic Nerve
-> Lateral Geniculate Nucleus
-> Visual Receiving Area of Cortex

light -> cornea -> aqueous humor -> pupil
-> lens -> vitreous humor -> retina
-> photoreceptor




## Accomodation

Ability to change focal length of lens by changing the curvature of the eye lens -> allows eye to adjust focus to different ranges (far/near)

Normal: Emmetropia -> clear vision at all distances -> image gets projected onto retina plane

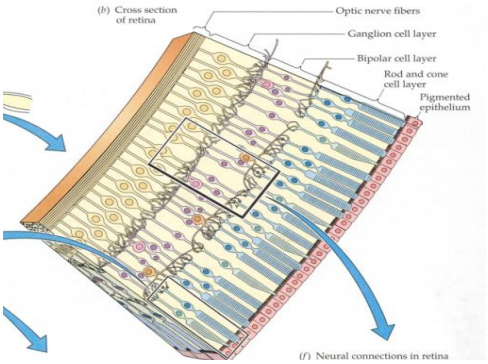
Far-Sighted: Hyperopia -> blurry: near -> inadequate refractive power, slower axial length -> image behind

Near-Sighted: Myopia -> blurry:far -> excessive ref. power, longer axial len. -> image in front

Age: Presbybopia -> similar to hyperopia -> insufficient accomodation due to aging: elasticity worse, --> lens get harder, ciliary muscle less power

- 15 : 10y, 10 : 20y, smaller than 2.5: 50+y

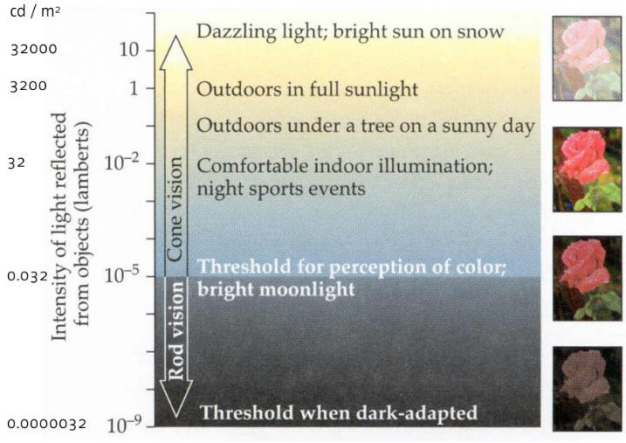
## Retinal Processing

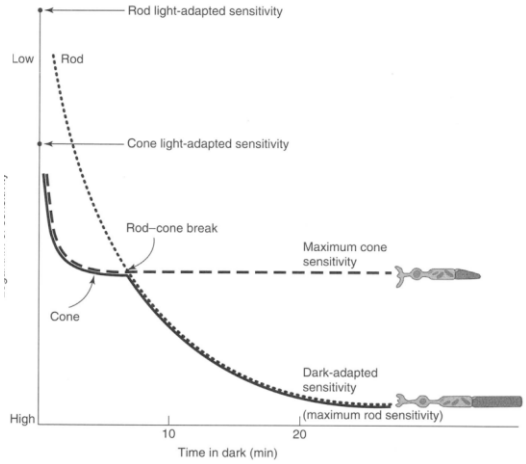


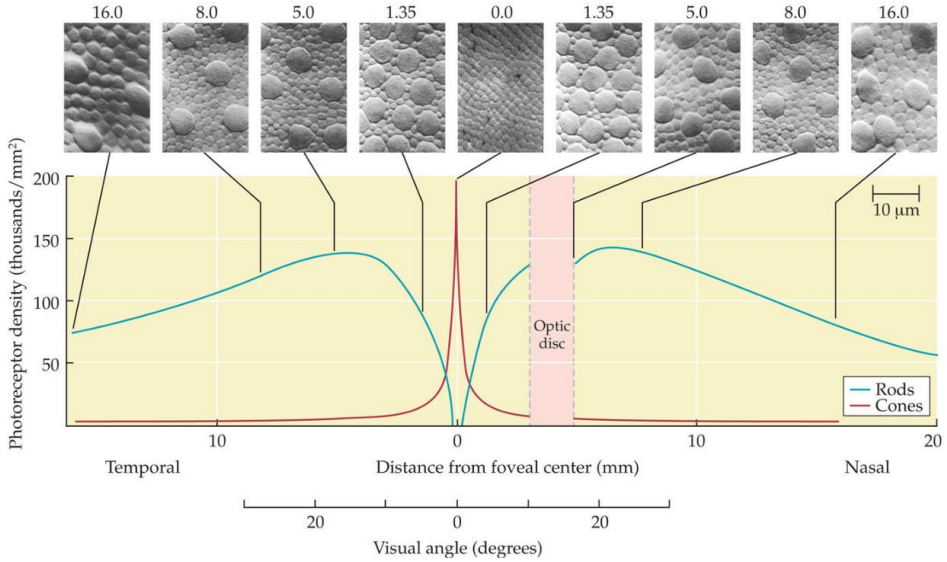

- Ganglion cells conduct electrical signals from eye to brain Network of vertical (bipolar) and horizontal connections between receptors and ganglion cells
- Cones and rods project onto the same ganglion cells
- Light path: Light -> Ganglion -> Bipolar -> Rod&Cone -> Photoreceptors *(light has to travel thru cells)*
- Blind Spot: Site where axons of Ganglion-Cells leave eye has no photoreceptors

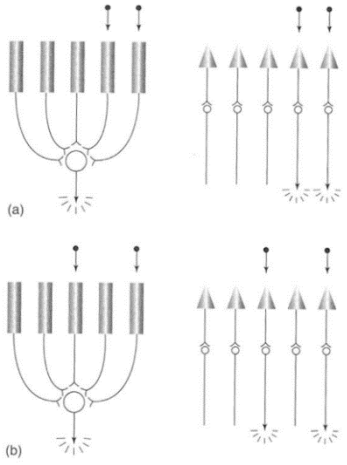
## Cone & Rod Vision

Cone: Color , Rod: Light




Rods: R+R+R+R = S -> S >=Treshold -> Response

Cones: C=S >= Treshold -> Response, C=S >= Treshold -> Response, ...

⇨ Summation increases (Light) Sensitivity but worsens Visual Acuity

| Feature | Photopic | Scotopic |
|---|---|---|
| Receptors | Cones (4-5 million) | Rods (90-100 million) |
| Photopigment | Three different cone opsins | Rhodopsin |
| Light sensitivity | Low, for day vision | High, also useful at night |
| Retinal location | Concentrated in the fovea | Outside the fovea |
| Acuity (highest resolution) | Very good in the fovea, lower in the periphery | Low |

**Properties of the fovea and periphery in human vision**

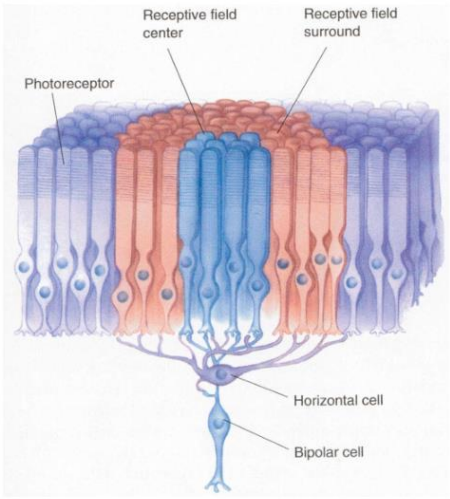
| Property | Fovea | Periphery |
|---|---|---|
| Photoreceptor type | Mostly cones | Mostly rods |
| Bipolar cell type | Midget | Diffuse |
| Convergence | Low | High |
| Receptive-field size | Small | Large |
| Acuity (detail) | High | Low |
| Light sensitivity | Low | High |

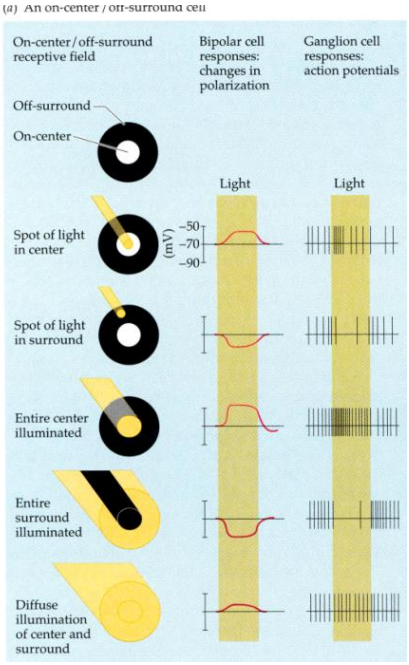
## Visual Acuity

- Quantification of sharpest point of vision
- sinusoidal gratings: distance between two cones in retina is about half an angular minute => a lighter and darker stripe must therefore fall on different receptors to see gaps
- VA depens on density of cones => this decays with retinal eccentricity
- rod system relatively constant: VA about 25 arcmin across FOV

## Retinal Ganglion Cell, Receptive Fields

**Lateral Inhibition**




two functional types of G-Cells sending projections to cortex

4 light sensitive cells: rods, 3 types of cones

# 03 Psychophysics (Methods)

**Sensation**: The ability to detect a stimulus and, perhaps, to turn that detection into a private experience / how our senses transduce energy from the world (light, sound, mechanical pressure) into neural energy

**Perception**: The interpretation of sensations and the assignment of meaning to them

**Birth of Psychophysics 1800, Fechner:** The science of defining quantitative relationships between physical and psychological (subjective) events—"physics on the x-axis and psycho on the y-axis …"

## Experimental Methods

**Thresholds**: Finding the limits of what can be perceived

**Scaling**: Measuring private experience

**Signal detection theory**: A statistical framework to understand how threshold-style decisions are made

**Sensory neuroscience**: The biology of sensation and perception

**Neuroimaging**: An image of the brain (mind?)

## Pyschophysical Methods

**Method of limits**: The magnitude of a single stimulus or the difference between two stimuli is varied incrementally until the participant responds differently.

**Method of adjustment**: Similar to the method of limits, but the participant controls the stimulus directly.

**Method of constant stimuli**: Many stimuli, ranging from rarely to almost always perceivable, are presented one at a time.

**Adaptive methods**: Latest development; here an algorithm selects the next presentation intensity based on the intensity of the stimulus and the response history of the subject. At least several dozen variants exist, both non-parametric ("up-down methods") as well as parametric (typically Bayesian) ones.

**Magnitude estimation**: The participant assigns values according to perceived magnitudes of the stimuli. (The previous four methods measure JNDs; magnitude estimation or scaling experiments attempt to directly measure the intensity of the "private" experience.)

**Method of triads with ordinal embedding**: View three stimuli and select two that are "most similar" to one another or, alternatively, select the "odd-one-out". Then try and find a (multi-dimensional) space in which to embed all the stimuli such that the distances in the space reflect the similarity judgements.#

**Cross-Modality-Matching**: Participant matches intensitiy of a sensation in (1) sensory modality with the intensity of a sensation (2) in another sensory modality

- JND-style measurements: most precise, reliable and immune to extranous influences
- Method of Constant Stimulus: preferable to adaptive procedures
- Rating, Scaling Experiments: much less reliable and immune, important to provide an anchor...
- Method of Triads: appears to be a good way to measure sensation

## Psychophysical „Laws“

**Two-point threshold**: The minimum distance at which two stimuli (e.g., two simultaneous touches) can be distinguished.

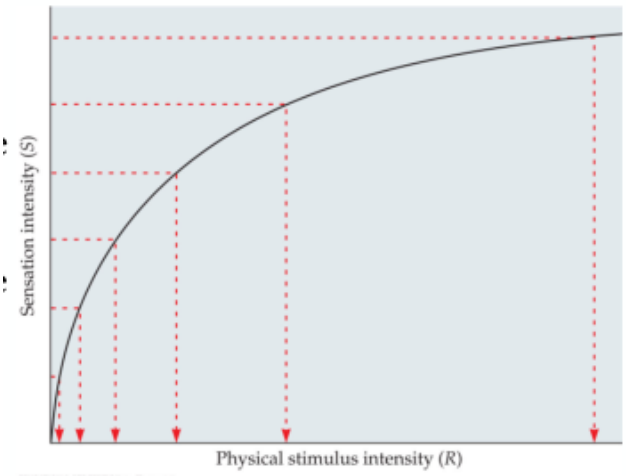
**Just noticeable difference (JND):** The smallest detectable difference between two stimuli, or the minimum change in a stimulus that can be correctly judged as different from a reference stimulus; also known as difference threshold.

**Absolute threshold**: Minimum amount of stimulation necessary for a person to detect a stimulus (typically 50% of the time).

### Weber’s Law, 1846

describes the relationship between a stimulus and its resulting sensation by proposing that the JND is a constant fraction of the stimulus intensity -> Thus, larger stimulus values have larger JNDs and smaller stimulus values have smaller JND

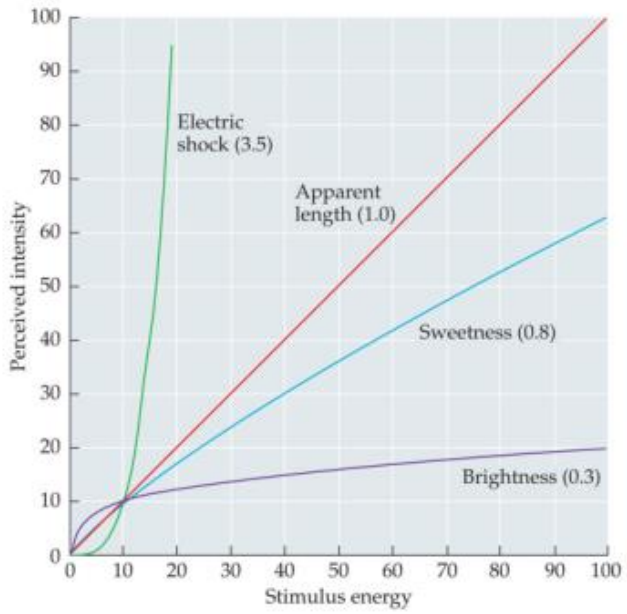
### Weber-Fechner-Law, 1860




A principle describing the relationship between stimulus magnitude and resulting sensation magnitude such that the magnitude of subjective sensation increases proportionally to the logarithm of the stimulus intensity

$$S = k \log(I)$$

### Steven’s Law, 1956




Magnitude Estimates: $S = a I^b$

Sensation (S) is related to a (possibly scaled) stimulus intensity (I) by an exponent (b)

## Signal Detection Theory, Thresholds

A psychophysical theory that quantifies the response of an observer to the presentation of a signal in the presence of (inevitable) internal noise.

**Internal noise**: Assumption that even in the absence of any external stimulus there is (variable) internal activity in the nervous system (c.f. the spontanous, non-zero firing rates of all neurones ever measured in the nervous system of any animal).

**Decision-axis**: For binary problems there is always a one-dimensional sufficient statistic independent of the dimensionality of the observation space. Often applied in a simple binary (or "diagnostic") setting with only four possible outcomes:

**Hit**: Stimulus is present and observer responds "Yes."

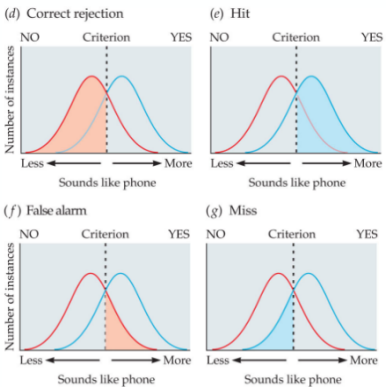
**Miss**: Stimulus is present and observer responds "No."

**False alarm**: Stimulus is not present and observer responds "Yes."

**Correct rejection**: Stimulus is not present and observer responds "No."

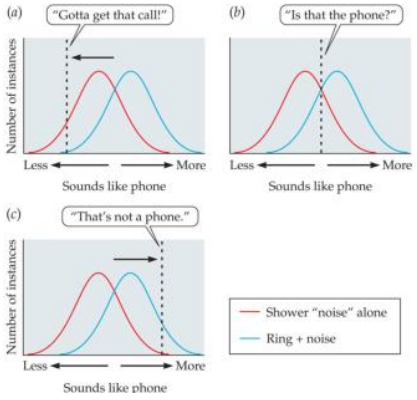
## Sensitivity vs. Criterion

Signal detection theory makes a distinction between an observer's ability to perceive a stimulus (sensitivity) and their willingness to report it (criterion). Sensitivity is the ability to detect the presence of a signal among noise, regardless of any response bias of the observer. The criterion, on the other hand, formalizes the idea of bias and indicates how willing the observer is to say "yes" to an ambiguous stimulus. An observer's criterion determines what kinds of errors they will make—whether there will be more false alarms or misses.




**Sensitivity**: A value that defines the ease with which an observer can tell the difference between the presence and absence of a stimulus or the difference between stimuli.

**Criterion**: An internal "threshold" that is set by the observer (note: decision, not sensory). If the internal response is above criterion, the observer gives one response. Below criterion, the observer gives another response



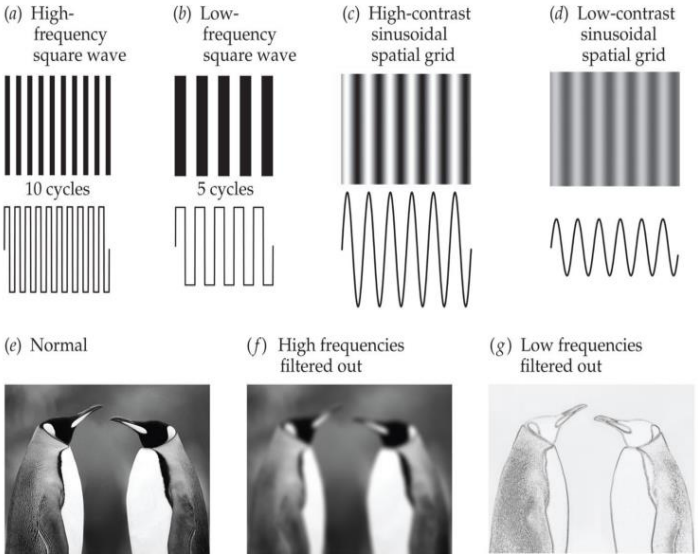

- ➔ Theory predicting how and when we detect the presence of a faint stimulus (signal) amid background stimulation (noise)
- ➔ assumes there is no signal absolute threshold and that detection depends partly on a person's experience, expectations, motivation and alertness

# 04 Spatial Vision

## Fourier Theory Reminder

Sine waves from a basis for (practically) any signal: Thus we can redescribe any signal—be it a sound, or an image—as the linear superposition of many sine waves.

### Properties of sine waves:




**Period or wavelength**: The time or space required for one cycle of a repeating waveform.

**Phase**: In vision, the relative position of a grating; in hearing, the relative timing.

**Amplitude**: The height of a sine wave, from peak to trough, indicating the amount of energy in the signal( corresponding to contrast in vision, loudness in hearing).

higher amplitude, higher contrast

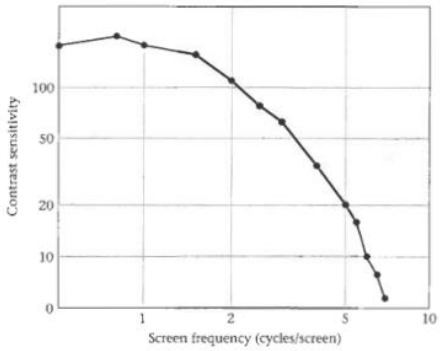
lower amplitude, lower contrast

Fourier analysis is a mathematical procedure for decomposing a complex signal into its component sine waves. **If the individual sine waves are re-combined, they will reproduce the original signal.** Fourier analysis is used extensively by perception researchers because it **provides a good description of stimuli** and also because **several perceptual systems perform Fourier analysis when processing stimuli** (e.g., the visual and auditory systems). In terms of stimuli based on sine waves, psychophysicists tend to use **pure tones in the auditory domain** and **sine wave gratings in the visual domain**. Sine waves may vary in their wavelength (distance for one full cycle of oscillation of the wave), period (time for one full cycle of oscillation of the wave), phase (relative shift of the sine wave) and amplitude (height of the sine wave, i.e. contrast in vision and loudness in hearing).

contrast sens. of retinal ganglion cells – Enroth-Cugell, Robson:

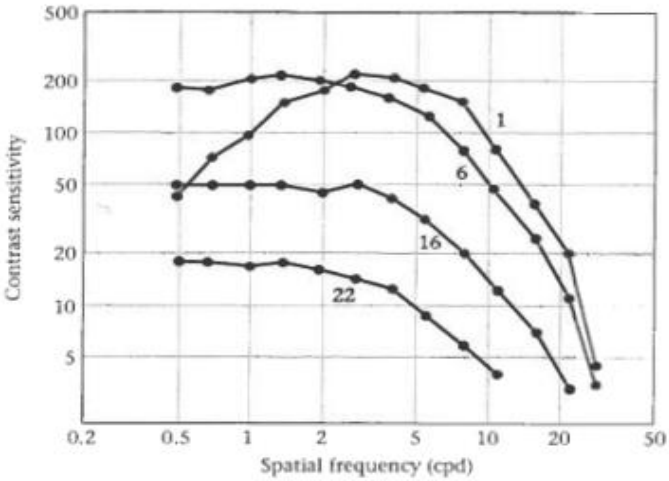
- supports: sensitivities of antagonistic centre and surround summating regions of ganglion cell receptive field fall off as gauss. func. of distance from field centre

## Contrast Sensitivty Function




-**fall off in (S)ensitivity** as cpd of test pattern increases due to **several components in Visual System being insensitive to high-spatial-freq**. targets: optical blurring reduces contrast, retinal ganglion cells (center-surround RF) less sensitive
**-no improvement of S at low spatial freq**.
**-small loss of S at lowest spatial freq.** but optical system does not reduce S therefore neural factors are at play (perhaps center-surround RF)
-high CS means: less contrast needed to see wave  pattern recognition

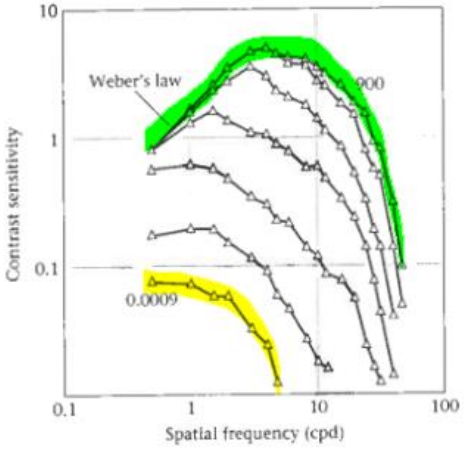
## CSF depends on temporal frequency




7.4 TEMPORAL VARIATIONS CHANGE THE SHAPE OF THE HUMAN SPATIAL CONTRAST-SENSITIVITY FUNCTION. The contrast-sensitivity functions shown here were measured with contrast-reversing targets at several different temporal frequencies. At low temporal frequencies (1 Hz) the contrast-sensitivity function is bandpass. At high temporal frequencies (22 Hz) the function is lowpass. Source: Robson, 1966.

... S falls at low freq., when measured with low temp. freq. (1Hz). At high temp. freq. (e.g during a series of rapid eye movements), there is no low freq. loss  images: flickered or slid over

## CSF depends on adaption or light level




7.21 HUMAN CONTRAST SENSITIVITY VARIES WITH MEAN FIELD LUMINANCE. Each curve shows a contrast-sensitivity function at a different mean field luminance level ranging from $9 \times 10^{-4}$ Trolands to $9 \times 10^{2}$ Trolands, increasing by a factor of ten from curve to curve. The stimulus consisted of monochromatic light at 525 nm. At the lowest level, under scotopic conditions, the contrast-sensitivity function is lowpass and peaks near 1 cpd. On intense photopic backgrounds the curve is bandpass and peaks near 8 cpd. Above these mean background levels, the contrast-sensitivity function remains constant. Source: van Nes and Bouman, 1967.

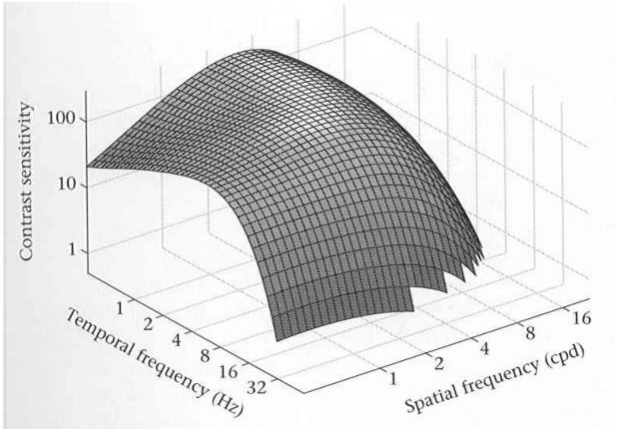
Low-Light -> rods activated -> observer must spatially average light signal to obtain signal -> cannot resolve high spatial freq.

⇨ scotopic: poor S to high cpd
⇨ low-pass

Normal-Light -> cones -> observer integrates over smaller spatial regions -> incr. spatial resolution

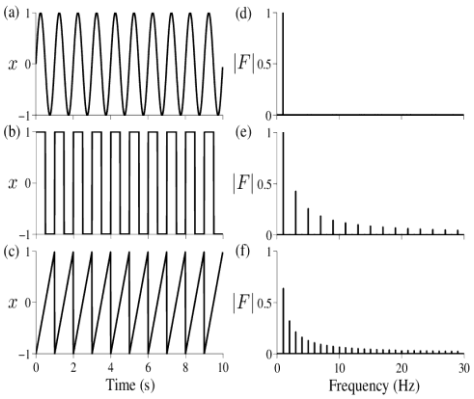
⇨ photopic: good S to mid cpd
⇨ bandpass

## CSF Surface




Allows us to predict the detectability of any stimulus—and for a wide variety of stimuli this is true

## Advent Modern Spatial Vision




**All visual stimuli can be represented by the sum** of (infinitely many) **sinusoidal gratings** *of different frequency, orientation, phase and amplitude*

if Visual System was SILS then its **response to any stimulus can be predicted from its response to sinusoidal gratings**, because they are the eigenfunctions of a SILS

To predict visibility of a stimulus we **consider its Fourier spectrum** rather than its pixel values (as **sinusoids are "dirac impulses" in the Fourier domain**)

## Michelson Contrast

$$c = \frac{L_{max} - L_{min}}{L_{max} + L_{min}} = \frac{L_{max} - L_{min}}{2L_{mean}}$$

## Campbell & Robson, 1968 or 67: Fourier Analysis To Visibility Of Gratings

1. The contrast thresholds of a variety of grating patterns have been measured over a wide range of spatial frequencies.
2. Contrast thresholds for the detection of gratings whose luminance profiles are sine, square, rectangular or saw-tooth waves can be simply related using Fourier theory.
3. Over a wide range of spatial frequencies the contrast threshold of a grating is determined only by the amplitude of the fundamental Fourier component of its wave form.
4. Gratings of complex wave form cannot be distinguished from sine-wave gratings until their contrast has been raised to a level at which the higher harmonic components reach their independent threshold.
5. These findings can be explained by the existence within the nervous system of linearly operating independent mechanisms selectively sensitive to limited ranges of spatial frequencies.

**<u>Detectability</u> of periodic patterns can be predicted from their Fourier spectrum.**

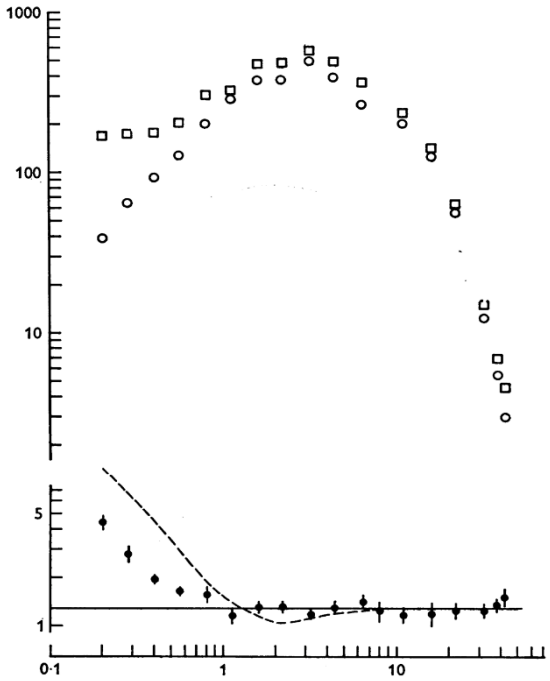
**DETECTABILITY -** Wandell_1995_ch7, p. 12

A look at Square-Waves:

- can be expressed as weighted sum of sinusoidal components using discrete fourier series
- square wave sq(x) with frequency f:
  - to get odd-number we use (2n+1) = odd
  - m being the given Amplitude of the square wave
  - sq(x) = m* 4/Pi * SUM[ ... * sin( ... * odd * f * x ) ]
- square wave at frequency f equals the sum of a series of sinusoids at odd numbered freq. -> f, 3f, 5f, ...
  - Amplitudes (= Contrast):
    - f: has Amplitude A ==> 3f: A/3, 5f: A/5
    - Amplitude very low then 7f, 9f, 11f, etc. are really small and can be ignored
  - to achieve appearance of a low-amplitude sq(x) it is okay to use only a few components (f, 3f, 5f):
    - $$sq(x) \approx \frac{4}{\pi}\left[\sin(2\pi f x) + \frac{1}{3}\sin(2\pi(3f)x) + \frac{1}{5}\sin(2\pi(5f)x)\right]$$
  - lowest freq. component is called **fundamental -> f**
    - has the highest amplitude -> A = 4/Pi therefore contains largest contrast information (?)
  - 3f, 5f, 7f, ... are harmonics (3f = 3rd harmonic, 5f = 5th harmonic, ...)

C&R use sq(x) to test multiresolution hypothesis (multi-channels):

- first measured smallest contrast level where square-wave-gratings detectable (JND --> contrast threshold for given square wave)
- argued: **neurons whoose receptive-field (RF) size is matched** to the **fundamental f will signal the presence of a square wave first**, making it the most important term in defining visibility of sq(x)
- if this is the case then **the threshold contrast of sq(x) should be 4/Pi times the threshold contrast of a sinusoidal grating** at the same frequency f




➔ the fundamental just being a sine function with an Amplitude of 4/Pi * m and that equalling contrast, square waves are inherently contrast-richer than sinusoids

What is plotted for the open symbols are the **contrast sensitivities** (1 over detection threshold) on the **y-axis** against **spatial frequency in cpd** . The **open squares** show the data for the **square-wave grating** , the **open circles** for the **sine-wave grating** . The **filled black circles** show the **ratio of the square-to-sine sensitivities** . The **solid black line marks the prediction at 4/pi** derived from the Fourier series of the stimuli. The dashed line marks the prediction of a simple peak-detector model of early spatial vision.

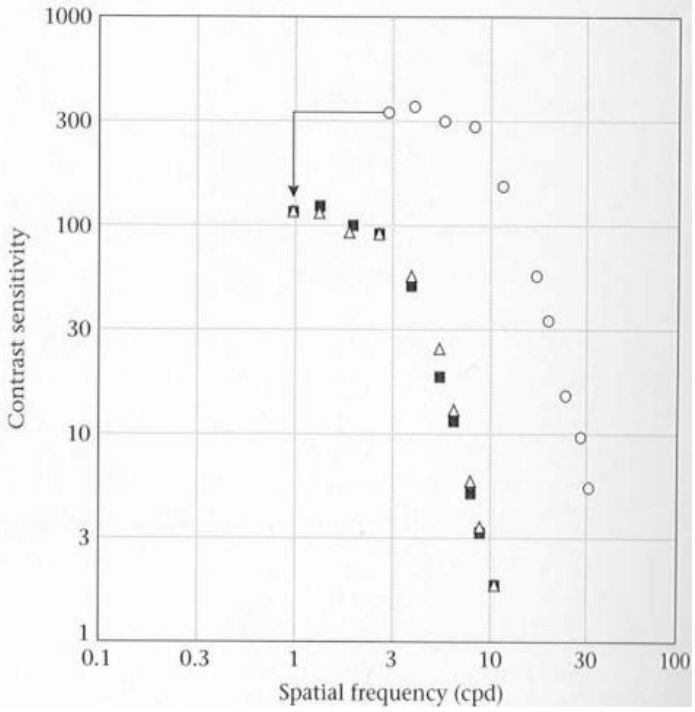
**DISCRIMINATION** - Wandell_1995_ch7, p. 12
perceived conrast stayed same so:

- sinusoid and sq(x) were held constantly on the same ratio therefore their detection rate was the same
- observer cannot distinguish square wave to sinusoid based on the fundamental component anymore
    - instead the higher-order components 3f, 5f, 7f, 11f,... come in play
    - recall: the components each have their own amplitude and the higher the amplitude of the fundamental, the higher they will be for the components, too

$$\text{sq}(x) \approx \frac{4}{\pi}\left[\sin(2\pi f x) + \frac{1}{3}\sin(2\pi(3f)x) + \frac{1}{5}\sin(2\pi(5f)x)\right]$$

    - so if the highest amplitude (= contrast perception) was in our fundamental but now we cannot use this anymore, the next in line would be the 3rd harmonic: 3f
    - therefore when 3f has a high enough amplitude the observer can distinguish the sq(x) from the sinusoid

Campbell and Robson found that observers discriminated between the sinusoid and the square wave when the contrast in the third harmonic reached its own threshold level. Their conclusions are based on the measurements shown in Figure 7.14. The filled circles show the contrast-sensitivity function. The open circles show the contrast of the square wave when it is just discriminable from the sinusoid. Evidently, the square-wave contrast needed to discriminate the two patterns exceeds the contrast needed to detect the square wave. However, we can explain the increased contrast by considering the contrast in the $3f$ component of the square wave. Recall that this component has one-third the contrast of the square wave. By shifting the square-wave discrimination data (open circles) to the left by a factor of 3 for spatial frequency, and downward by a factor of 3 for contrast, we compensate for these two factors. The open triangles show the open circles shifted in this way. The open triangles align with the original contrast-sensitivity measurements. From the alignment of the shifted discrimination data with the contrast-sensitivity measurements, we can conclude that the square wave can be discriminated from the sinusoid when the $3f$ component is visible at detection threshold.




**7.14 DISCRIMINATION OF SINUSOIDAL AND SQUARE-WAVE GRATINGS** becomes possible when the third harmonic in the square wave reaches its own independent threshold. The open circles plot the contrast-sensitivity function. The open triangles show the contrast level at which a square wave can be discriminated from its fundamental frequency. The filled squares show the square-wave discrimination data shifted by a factor of 3 in both frequency and contrast. The alignment of the shifted curve with the contrast-sensitivity function suggests that square waves are discriminated when the third harmonic reaches its own threshold level. Source: Campbell and Robson, 1968.

## Detection of Grating Patterns: Single and Multi-Channel-Models, Graham & Nachmias, 1971

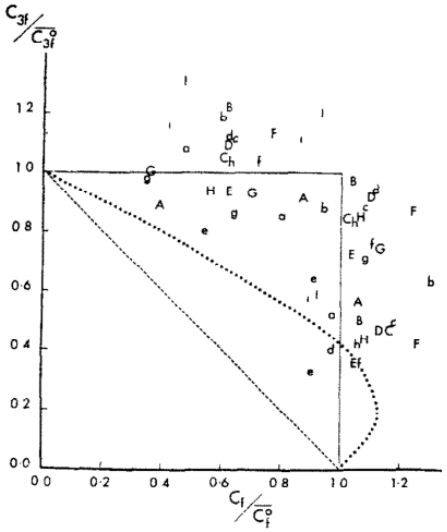
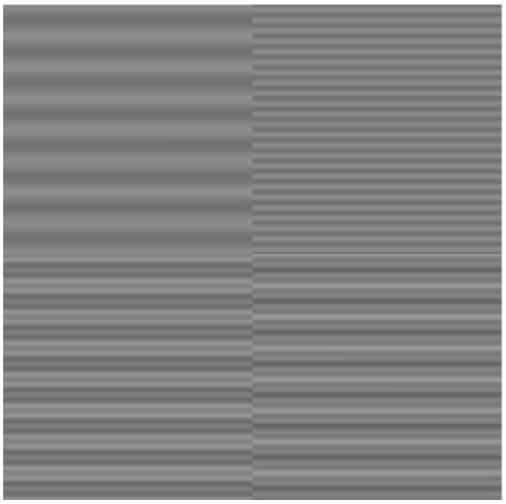



FIG. 3. Contrast thresholds for complex grating patterns containing two frequencies, *f* and 3*f*, in two different phases. The coordintes are the contrast of each component in the complex grating at threshold relative to the threshold contrast of the corresponding simple grating. Results obtained with the peaks-add form of complex gratings and with the peaks-subtract form are plotted as capital and small letters, respectively. The corresponding predictions of the single-channel model are represented by the diagonal dashed line and the dotted curve. The upper and right edges of the square represent the predictions of the multiple-channels model for both peaks-add and peaks-subtract gratings.

Detection of compound patterns with sufficiently different spatial frequency is independent of local phase ("summation experiments")

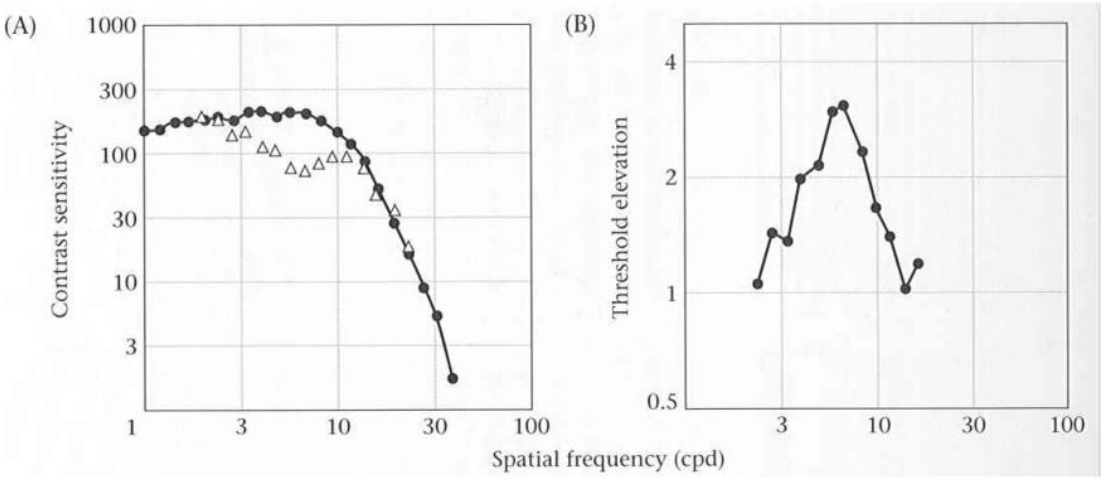
This means, regardless of local phase (resulting in higher summation) the perceived contrast stays the same



## On the existence of neurons in HVS selectively..., Blakemore&Campbell, 1969

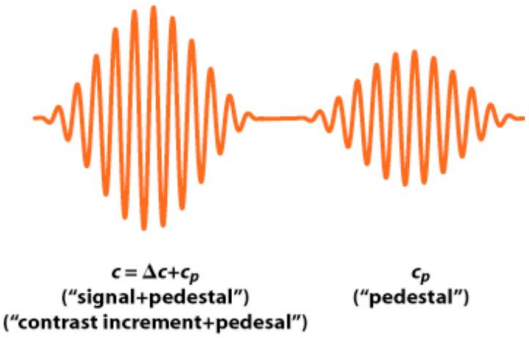
Adaptation is spatial frequency selective

- ➔ look at specific contrast freq. for long time -> look at different contrast freq. -> will not see, but will see different pic of same contrast freq. => adaption
- ➔ moving pattern will increase sens.
- ➔ neuron selectively sensitive to a spatial freq. -> due to orientation and interocular transfer of adaption, these neurons probably are in visual cortex
- ➔ recognition of complex images & generalization for magnification




**Multi-Resoultion-Theory: CSF as the sum of more narrowly tuned elements (channels)**

## Contrast Discrimination may be better than detection, Nachmias & Sansbury, 1974



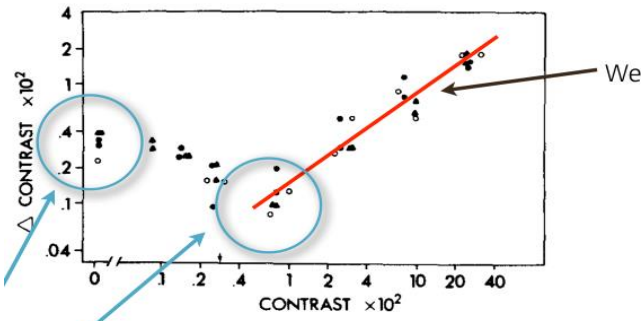
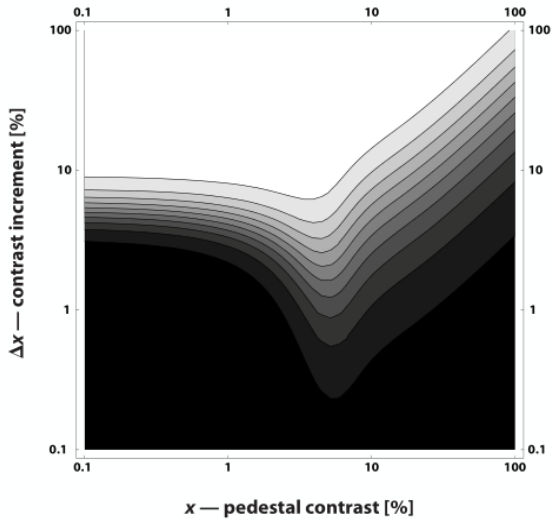




Fig. 1. The just discriminable contrast difference between a pair of 3 c/deg gratings plotted against the lesser contrast on log–log coordinates. Different symbols are used for different observers. The arrow on the abscissa points to the average value of the absolute contrast threshold.

**Contrast Discrimination Model: Dipper Func.**




Dip -> non-linearity

Low contrast -> accelaration (amplification)

Threshold reached -> deaccel. -> Linear -> Webers Law

## Uncertainty Experiments, Davis, Kramer & Graham, 1983

Prevalent theories of pattern vision postulate mechanisms selectively sensitive to spatial frequency and position but not to contrast. Decreased performance in the detection of visual stimuli was found when the observer was uncertain about the spatial frequency or spatial position of a patch of sinusoidal grating but not when he was uncertain about contrast. The uncertainty effects were consistent with multiple-band models in which the observer is able to monitor perfectly all relevant mechanisms. Performance deteriorates when the observer must monitor more mechanisms, because these mechanisms are noisy and give rise to false alarms. This consistency is further evidence that the spatial-frequency and spatial-position mechanisms are noisy, a conclusion previously suggested by the "probability summation" demonstrated in the thresholds for compound stimuli. Somewhat paradoxically, the Quick pooling model, which quantitatively accounts for the amount of probability summation in pattern thresholds, predicts no effects of uncertainty. It cannot, therefore, be strictly correct.

Uncertainty about spatial frequency, spatial position, or contrast of visual patterns

## Three Key Findings for Modern Spatial Vision

**Detectability** (treshold measurements) & **Discriminability** (supra-threshold measurements) of **simple gratings** (sine, square, sawtooth) can be predicted from **knowledge of CSF** and the **Fourier Spectrum of the gratings**

**Adaption** to gratings of certain frequency, **only affects perception of gratings with similar freqeuncy**

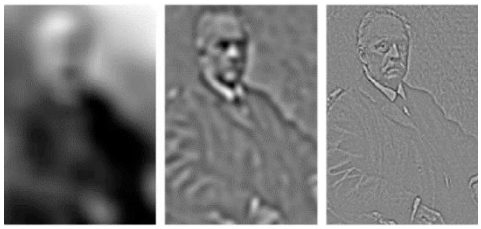
-> this all suggests or is **consistent with the Multiresolution** T**heory**

- system is **tuned to narrow ranges of spatial frequency**
- **CSF is a sum of many such channels**

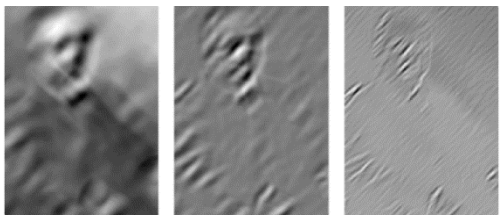
# 06 Early Spatial Vision

**Prior** to the seminar work of **Campbell and Robson** published in 1968 , researchers in pattern perception, often referred to as **early spatial vision/spatial vision**, **thought of the stimuli exclusively in the space domain, in terms of lines, corners and edges**. **After** the publication of "Application of Fourier Analysis to the Visibility of Gratings" in the Journal of Physiology, however, vision researchers up to this day always consider **stimuli** also in the **Fourier domain**. Additional experimental **data consistent with the linear, independent multi-channel model** came, e.g., from Blakemore and Campbell's adaptation studies, or from the famous *1f, 3f and phase manipulation* experiments by Graham and Nachmias, published in 1971, or the elegant experiment by Carter and Henning from 1971 , showing that a single cycle of a sine-wave grating could be easier to detect than many cycles if the signal was masked by narrow-band visual noise. Whilst there exists a **large body of work supporting the linear, independent multi-channel model**, there are **notable exceptions**. One of the most prominent is a study by Henning and colleagues from 1975, based on an auditory phenomenon, the "missing fundamental".

Helmholtz -> several filters:



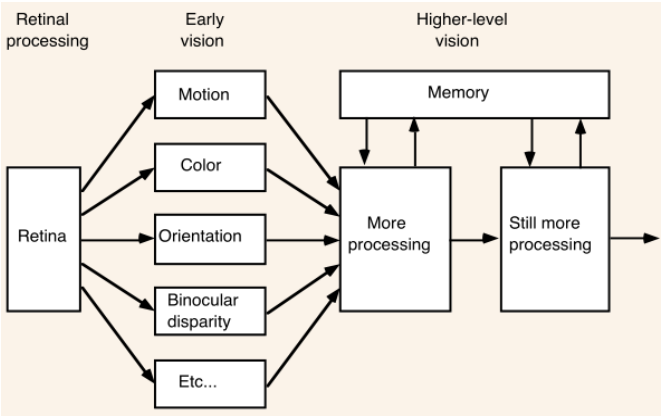
*2 circularly-symmetric bandpass filters = neural images of retinal ganglion and LGN cells*



*1 oriented bandpass filters (45˚ example shown) = neural images of V1 simple cells*

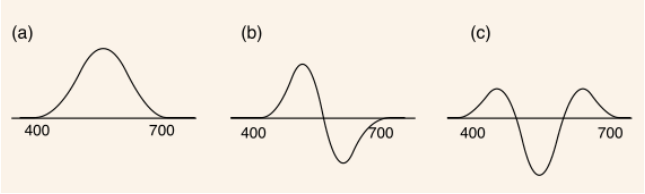
Detection, summation, adaption experiments consistent with the **early stages of the visual system behaving like linear, independent filters of limited bandwidth**, which has striking similarity between Helmholtz's theory of frequency representation in the auditory system and Campbell & Robson's model of visual spatial frequency analysis

## What are the fundamental substances of vision (what is early vision)?




At this level it is premature to talk about discrete objects, even such simple ones as edges and corners. There is general agreement that early vision involves measurements of a number of basic image properties including orientation, color, motion, and so on… the first stage of processing involves a set of parallel pathways  each corresponding to one particular visual property  these are the elements of early vision

Local avg. of derivative = derivative of local avg.




a: Gaussian Func. G(x) = Luminance
b: G'(x) = Blue-Yellow
c: G''(x) = Red-Green

low-order derivatives = 2D-Receptive Field Types

## Evidence for the Standard Early Spatial Vision Model

- Detection & discrimination of simple patterns from Fourier Spectrum of stimuli *(C&R,1986)*
- Detection of Compound Gratings  independent Multi-Channel *(G&N, 1971)*
- Adaption Studies -> Multi-Channel *(B&C, 1969)*
- Prediction of contrast discr. Behaviour given within-channel non-linear transducer estimated from data *(Foley & Legge, 1981)*
- Band-limited channels = local derivative operators *(Adelson & Bergen, 1991)*

## Redundancy in natural images

- Experiment: Barlow
- Image with salt&pepper noise  ask participants to reconstruct original image  highly accurate results  we take advantage of the spatial redundancy in img data

## Summary of natural image statistics

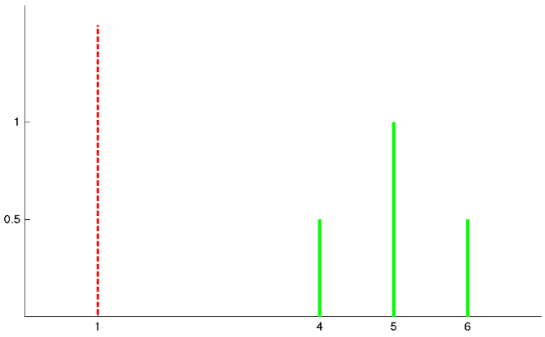
- Images are highly redundant  can predict individual pixel from neighbouring pixels, strong statistical regularities
- These statistical regularities are so that the most independent basis functions for images look similar to visual cells in early visual areas (Hyvärinen&Oja,2000)
- Spatial-frequency tuning of visual cells (“zero-mean bandpass-filter”) / multi-channels are redundancy reduction

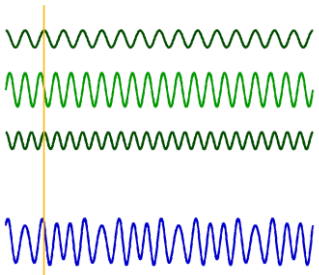
## Contradicting Experiments to Standard Model

*Henning, Hertz & Broadbent, 1974:*

Bruce Henning and colleagues published a series of **experiments** in the mid 1970s which were **inconsistent with the independent multi-channel model of Campbell & Robson**. Henning et al.'s experiments were **inspired by the "missing fundamental" in auditory pitch perception**, and they used both **amplitude modulated (AM)** as well as **quasi-frequency modulated (QFM) gratings as stimuli**.

*shows the amplitude spectrum of both AM and QFM gratings*



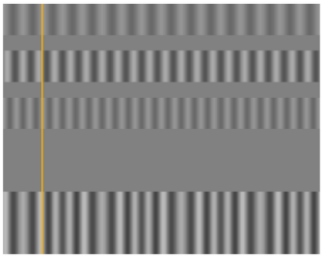
*shows the appearance and the cross-section through a QFM grating and its constituent 4f, 5f and 6f gratings*

Henning et al. reported to find **strong interactions (masking) between a sine-wave with frequency 1f and an AM grating composed of 4f, 5f and 6f**:

<u>According to Campbell and Robson there should have been no interaction (masking) between the stimuli</u>, however. Furthermore, **there was clearly less masking with a QFM grating as opposed to the AM grating**, and <u>this should not have happened, pointing to the importance of phase relations between the stimulus components, contrary to the findings of Graham and Nachmias</u>

Abstract: **Gratings with three sinusoidal components of highspatial frequency are shown to interact with a sinusoidal grating two octaves lower in frequency.** This finding is inconsistent with the hypothesis that the visual system analyses spatial patterns in independent narrowly-tuned bands of spatial frequency.

There are more experiments that contradict the Standard Model by C&R  (this one) Henning 1975 and Derrington & Henning 1989



*shows the respective graphs for a AM grating*

# 07 Object Recognition 1 – mid-level vision

- Is fast -> as fast as 150 ms
    - So fast that there cannot be a lot of feedback from higher brain areas
- Is a Feed-forward process: process that carries out computation -> object recog. One neural step after another, without needing feedback from later stage to earlier stage
- Very flexible

## Problems of Perceiving and Recognising Objects

- Several pictures (just pixels on screen) shown and in each of them them perceived a house (different styles, angles, …)
- How are we able to recognise these different pictures as houses?
- How can we see that it's the same house form different angles

## Perceptual Organisation

- How does the visual system move from points of light (pixels) to whole entities in the world -> house
- Same information with different representations:
    - Bunch of value data as retinal input then processed in the visual system to a entity

## Mid-level Vision: Contours and Gestalt

Middle vision refers to a set **of processes that combine features detected in early vision (such as edges and contours) into objects**. Middle Vision **utilizes rules and principles for combining elements into perceptual groups**, which were **discovered by psychologists from Gestalt tradition**. Some important steps here were*: finding edges of objects, dealing with occlusion, texture segmentation and grouping, detmining figure/ground assignments.*

➔ Perception of Edges and Surfaces
➔ Determines which regions of an image should be grouped together into objects

**Illusory Contours**

A contour that is perceived even though nothing changes from one side of the contour to the other. The image is constructed only from such contours but we perceive a house on top of filled out circles—though factually it's only some "Pac-Men" and disconnected lines. There are Rules to "see" a contour like good continuation.

Gestalt (form) grouping rules: a set of rules that describe when elements in an image will appear to group together like: two element will tend to group together if they lie on the same contour

## Early Approach to Perception: Gestalt Movement

**Structuralist Perspective**, *Wundt, Mach, v. Ehrenfels*, ~19. Jahrhundert

- Sensory atoms: primitive, indivisble elements of experience
- Perception from rapid, unconscious associations between sensory atoms (memory-like linking of experiences)
- Concatenation: Associations were simply added together  joining points on the retina
- Learn more about world via associations  perception becomes richter and complexer
- v. Ehrenfels: from the association of sense atoms we get extra element  Gestalt-Qualität
    - due to this, a tune can be transposed to a new key, using completely different notes while still retaining its identity
- predict recognition to be viewpont invariant but it has been shown to be dependent

Then came the **Gestalt Movement**, in rise against structuralism

- inverted structuralist theory: Gestalt-Qualität is the immediate perception which cannot be reduces to atoms
- *Max Wertheimer (1880-1943), Wolfgang Köhler (1887-1967) and Kurt Koffka (1886-1941)*
- *Wertheimer*: You hear the melody first and only then can it be divided into individual notes  similar in vision with dotted line

**Emergent Properties:**

- Squares have 3 perceptual properties: color, size, position
- Mutiple squares arranged will form new properties of length, orientation and curvature  Emerging

Perception of illusory contours depend on seemingly subtle manipulations of display

**Multistability:** Primary Evidence against sensory atoms being the essential part of perception, was "reversible" / "multistable" figures

Proof: List of all the individual sensations involved when seeing a vase would be the same as when seeing a face but the perception of both are very different to eachother => perception cannot be reconstructed one-to-one from sensory atoms

➔ **The perceptual whole is greater than the sum of its parts**

**Extending Gestalt Principles into a whole Theory:**

- *Köhler*: Chimpanzees can learn via „sudden insight" into the structure of the problem and not just incrementally via trial and error (*Pavloc & Thorndike*)
- Mental processes analogous to force fields in physics -> relationships with added parts
  - Field depends on structure of configuration
- *Koffka, 1935* in Principles of Gestalt Psychology: Basic Question of vision research
  - „Why do things look as they do?"
  - Integrate the facts of inanimate nature, life and mind into a single scientific structure

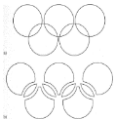
## Gestalt Principles of Perceptual Organisation

Based upon Perceptual grouping: how are elements of a complex display perceived as „going together"?

- **Proximity: nearby items**
- **Similarity: similar-looking items**
- **Common Fate: items that move together**
- **Symmetry: symmetrical item**
- **Parallelism: parallel items**
- **Continuity: smooth curvature**
- **Closure: closed figures**

### *Law of Prägnanz – Good Form / Good Gestalt*

Wertheimer tried unifying Gestalt Laws under one general principle -> Law of Prägnanz



The perceptual field will take on the simplest and most encompassing structure permitted by the given conditions  try to achieve maximum stability with minimum expenditure of energy

## Criticism of Gestalt Theory

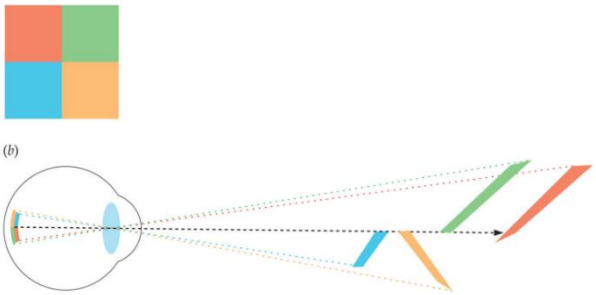
*Bruce, Green & Georgeson, 1996*

„The physiological theory of the gestaltists has fallen by the wayside, leaving us with a set of descriptive principles, but without a model of perceptual processing. Indeed, some of their "laws" of perceptual organisation today sound vague and inadequate. What is meant by a "good" or "simple" shape, for example?"

- not really laws like in science
- don't allow prediction of behaviour in novel circumstances
- „laws" are unclear if they independent to eachother, not based on experiments…

# 08 Object Recognition 2 – more on mid-level vision, neuroscience

## Accidental Viewports




A viewing position that produces some regularity in visual image that is not present in the world

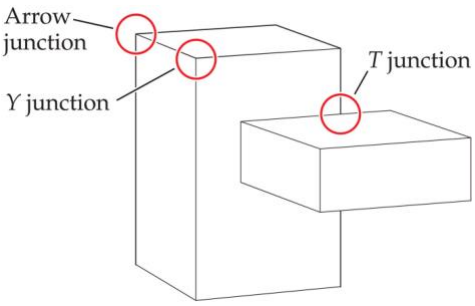
Visual system assumes viewpoints are not accidental

Forced Perspectives -> Reality Warping

## Non-accidental Features - the generalised-cone components / recognition-by-components

*Witkin,Tenenbaum, 1983*

A feature of an object that is not dependent on the exact (or accidental) viewing position of observer




**T-Junctions**: indicate occlusion, Top of T is in front and stem of T is in back

**Y-Junctions**: indicate corners facing observer

**Arrow-Junctions**: indicate corners facing away from observer

*Biedermann, 1987*

Properties of edges in 2D → visual system takes that as evidence → same properties in 3D world

- straight line in 2D (collinearity) → visual system infers: edge producing line in 3D also straight
- visual system ignores possibility that property in image might be result of a accidental alignment of eye and curved edge

**Properties**:

- Collinearity (of points or lines)
- Cuvilinearity of points of arcs
- Symmetry
- Parallel Curves
- Vertices

## Figure-ground

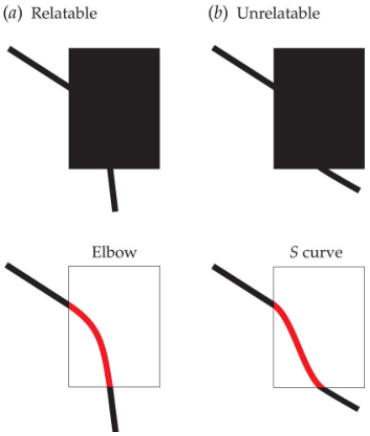
**Figure**: foreground object
**Ground**: background
**Assignment**:

- process of determining that some regions of an image belong to a foreground object
- other regions are part of background

**Gestalt figure-ground assignemnt:**

- **surroundedness**: surrounding region = ground
- **size**: smaller region = figure
- **symmetry**: a symmetrical region = figure
- **extremal edges**: edges of object shaded so they seem to recede in the distance = figure
- **releative motion**: one region moving in front of another -> closer region = figure

## Occlusion:




**Relatability:** the degree to which two line segments appear to be part of the same contour

Two edges are relatable if they can be connected with a smooth convex / smooth concave curve (a) but not if the connection requires an S curve (b)

## Parts and wholes:

*Navon, 1977*

**Global superiority effect**: properties of the whole object take precedence over the properties of parts of the object

⇨ we process global aspects of an image before local aspects

## Texture Segmentation and grouping:

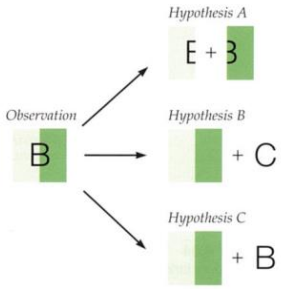
**Texture Segmentation**: carving an image into regions of common texture properties
**Texture grouping**: depends on statistics of textures in one region versus…
**Camouflage**: animals exploit gestalt grouping principles to group into their surroundings → sometimes used to confuse observer (war)

## Summary

Five Principles of Middle Vision

1. bringt together that which should be brought together
2. split asunder that which should be split asunder
3. use what you know
4. avoid accidents
5. seek consensus and avoid ambiguity

## Bayesian approaches to perception



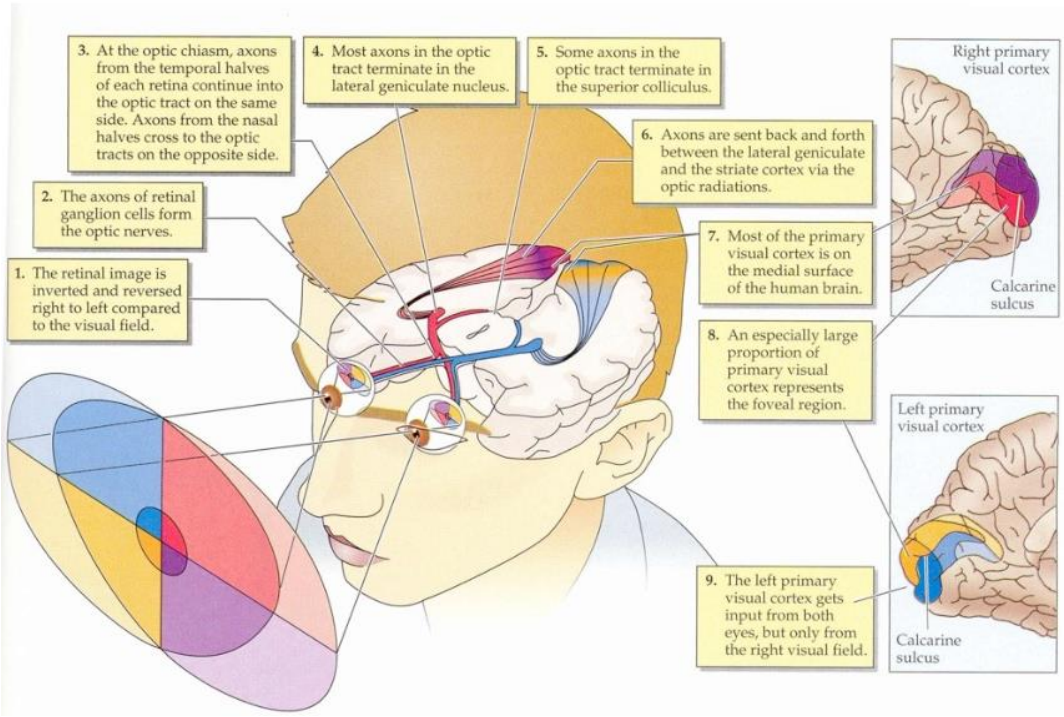

Prior: what you believe beforehand, knowledge of mid-level principles comes here

Likelihood: how consistent is the input with all possible hypotheses?

Posterior: combination of likelihood and prior

Stimulus → **visual system tries to figure out most likely situation that has produced this pattern of activity**

## Neuroscience of object recognition



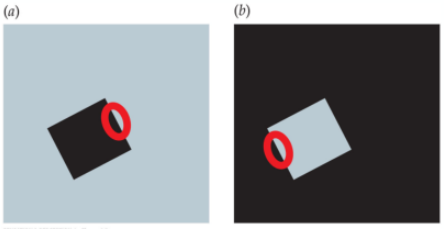

### From spots and bars to objects and space

how does brain recognise objects?

- Retinal ganglion cells = „spots“
- V1 = „bars“ (local, blurred, derivative operators – „stuff“)

➔ After V1 something quite sophisticated must happen

### Receptive fields in extrastriate areas




**Extrastriate cells**: more sophisticated than those in striate cortex → respond to more complex properties
„border ownership“: for a border which side is part of object which is part of background

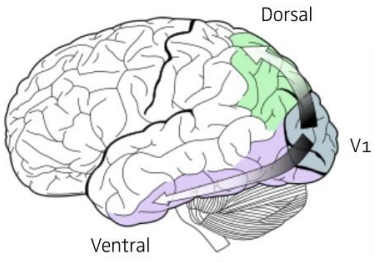
Same visual input in both (a) and (b) then a V1 neuron would respond equally to both

A V2 neruon might respond more to (a) than (b) due to the black edge being owned by the square in (a) but not in (b)

V2 also seems to be selective for texture: (*Freeman, Ziemba, Heeger, Simoncelli, Movshon, 2013*)

- Responding more to naturalistic texture than to amplitude-matched noise
- V1 doesn't differentiate the two

### Two Streams Hypothesis (= what and where pathways)




Visual Cortex organised into two relatively specialised processing streams:

- **What pathway (ventral stream):** relative specialisation for object recognition (shapes, names and functions)
- **Where pathway (dorsal stream):** relative specialisation for spatial localisation and guidance of action

Evidence: a lot of it in neuroanatomical, electrophysiological, lesion studies; earliest cues from studying humans with localised brain damage (lesions)

- **Lesion**: noun - a region of damaged brain, verb – to destroy a section of brain

### Lesions to Inferior Temporal

Inferior Temporal (IT) cortex lies at the end of ventral stream → if IT lesioned leads to **agnosias**
**(object) Agnosias**: Failure to recognise objects in spite of ability to see them (can copy image but not name it). Lesioning IT cortex (in monkeys) leads to visual deficits in discrimination and recognition analogous to agnosia in humans
**Prosopagnosia:** Inability to recognise faces

- Receptive fields in IT neurons are very large
- don't respond well to spots or lines but respond well to stimuli like hands, faces or objects (therefore if damaged leads to problems with recognition)

### Lesions to Dorsal Stream

**Optic ataxia**: inability to visually guide arm movements
**Hemispatial neglect**: patient is unaware of contralateral region of space
**Akinetopsia**: inability to perceive motion

### Double Dissociations → two-streams theory

- **Lesions to ventral stream damage object and shape perception** (IT) but not motion or spatial perception
    - if agnosias then can recognize faces but not objects
    - if prospagnosia then can recognize objects but not faces
- **Lesions to dorsal stream damage motion and spatial perception** (MT, MST) but not object recognition

### Functional Specialisation, cells very specialised to certain objects in the world

Single neuron for example responsible for recognizing grandmother (originally *Barlow, 1972*)

- Seems very unlikely → computation and "storage" wise quite inefficient → not robust
- Some believe this though → *Quiroga et. al, 2005*: found cell that responds specifically to Jennifer Aniston

### Caveats and outstanding questions

1. Most terms ill-defined ("parallel-processing, "feedback") – *Douglas & Martin, 1991*
2. Receptive versus projective fields (*Lehky & Sejnowski, 1988*)
3. Anatomy: cortical pathways not based on functional anatomy i.e know about number of fibres only not their importance or power
4. Do we really understand V1 in its fullest?

# 09 Object Recognition 3 – Object Representation

**What does "object recognition" mean?**

- Categorization of perceptual experiences, hierarchical problem
- *Aristotle*: categories are defined by necessary and sufficient conditions for membership
- *Wittgenstein, 1953*: there is no such set of necessary and sufficient features that apply to natural categories → instead used analogy of **family resemblance**:
    - Members of the same family look similar not because of particular features but due to global similarities (cannot be captured in simple logical rules)
- *Eleanor Rosch, 1973, 1975*: suggested **Prototypes-> the "best example" of each category**
    - Average member (doggiest dog) or ideal (reddest red)
    - This allows for graded category membership as distance to defined prototype
    - Study: people rate category members as good or not => **category levels**

## Category Levels – Rosch

- **Basic-Level category**
    - the highest level category in which members have similar shape, motor interaction and common attributes
    - intermediate level of hierarchy (dog|table)
- **Superordinate-level category**
    - More general term for an object (animal|furniture)
- **Subordinate-level category**
    - More specific term for an object (dalmation|dining table)
- **Entry-Level category**
    - The label that comes in mind most quickly when identifying object (irrelevant of subordinate or basic level)
- Jolicoeur, Gluck & Kosslyn, 1984: Basic level categories are not universly defined for the enite category

➔ **Object Recognition usually refers to the classification of objects into entry-level categories**

- **Object Recognition:** realising that you have seen a given object before, regardless of whether you can name object (-> object memory)
- **Object Identification:** recognising a particular known object ("my cat")
- **Object categorisation/classification:** classiyfing objects into entry-level categories

## Alternatives

### Recognition-By-Components (RBC) – structural model by Biederman 1987

**When an object is perceived it is represented as a series of volumentric parts (geons)** and the **categorical relations between parts (above, below, beside)**. Once an object is represented as volumentric parts and spatial relations the **process of object recognition itself is rather straightforward and invariant with viewpoint**

- Popularized by Marr, 1982
- Four-stage theory of vision: image-based, surface-based, object-based and category-based processing → generative system
- People know ca. 30.000 different categories → novel object recognition

**Geons**: geometric ions (basic shapes) of which objects are build

- Blocks, cylinders, wedges, cones

**Non-accidental Relations** (spatial relationship between geons)

- Edge, Symmetry, Size, Axis

### *Accounting for empirical phenomena*

Typically effects (Roschs Prototypes): prototypical instances (a robin) activate categorical representation (bird) more strongly than atypical instancs (ostrich)

Entry-Level categories: atypical examples (ostrich) activate their subordinate category (ostrich) more than their basic-level category (bird)

### *Limitations of RBC*

Based on concrete objects which have specified boundaries → chair -> 3 chairs but sand, water,snow cannot be numbered in that way ➔ suggests these are identified through surface characteristics (texture, color) rather than with volumetric primitives; Problems with non-rigid objects

### *Viewpoint Invariance*

1. A property of an object does not change when observer changes viewpoint
2. Class of Theories → RBC
3. Observation that humans are good at recognising 3D objects despite variations in perspective

But Object Recognition is not completely viewpoint invariant → objects in canonical views are classified and recognised faster & more accurately

### *Structural-description theories summarized*

Problems: canonical view effects, nonrigid objects (Wichmann disagrees) but geons are not the best for describing non-rigid deforming, non-geometric objects or natural shapes → anything not man-made

## View-Based Theories

objects are represented as collection of remembered views of the object (views are stored as templates). Therefore initial representation of the object is easy but matching the perceived view to representations in memory is difficult → object recognition should be slower for objects seen from novel view-points

DiCarlo, Zoccolan, Rust, 2012: …the ability to rapidly recognize objects despite substantial appearance variation, **is solved in the brain via a cascade of reflexive, largely feedforward computations that culminate in a powerful neuronal representation in the inferior temporal cortex**. However, the algorithm that produces this solution remains poorly understood.

- ⇨ Perhaps there are several object recognition processes depending on type of object
- ⇨ Perhaps view-based representation exist when current input matches stored view (fast and accurate recognition) but under challenging conditions categorisation relies on a slower process of matching structural descriptions

## Faces

Faces are different than other objects because all faces have the same parts in the same relationships with one another. Therefore, fine metric details of faces are important in recognition and it seems the visual system represents faces hollistically in terms of these fine metric details whereas it does not in the case of objects. … inverted faces hard to recognize…

## Summary

- Perception -> Feedback and Reentrant Processing
- Initial object recognition can be really fast (150ms)
- Brain continues to process information sending signals up and down “what-pathway”(ventral stream)
- ⇨ Object Recognition = conversation among many parts of brain

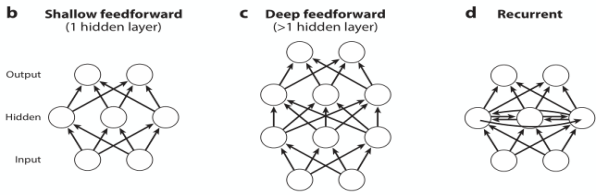
# 10 Object Recognition 04 – Algorithms, DNN

**DNN**: loose terminology to refer to networks with at least two hidden or intermediate layers, typically at least five to ten or twenty

- ImageNet challenge with 1000 categories and 1.2 million training images
  - more than 100 types of dogs → not possible for humans...
  - train and optimize your machine with training data
    → check if it's good with 50 000 test images
- AlexNet reduces prediction error by 50% → deep nets!! even better than average human (but categories not suitable for humans)

**Neural Networks**:

- 50s/60s ANN, perceptrons → computations by neurons
  - single layer → simple weighted combination of inputs
- 80s: back-propagation algorithm as learning rule for multilayer networks (connectionist networks)
  - 3layer with infin. hidden units is universal function approximatior, can compute everything tha is computable
    → BUT: lack of theory, non-convex optimization problems, little computing powers
- 2005: deep neural networks , atleast two hidden/intermediate layers
- increase of training data, computing power and tricks with simple non-linearities (ReLu) and convolutional rather than fuly connected (everything connected to everything) → method of choice in ML
- RELU: rectified linear, 0 till a threshold then linear




- recurrent network: not feed-forward, also connections between within a layer
- if it would be linear → results would be limited to plane (Fläche )
- complex forms

*Yamin et. al: Performance-optimized hierarchical models predict neural responses in higher visual cortex: The **ventral visual stream underlies key human visual object recognition abilities**. However, neural encoding in the higher areas of the ventral stream remains poorly understood. Here, we describe a **modeling approach that yields a quantitatively accurate model of inferior temporal (IT) cortex, the highest ventral cortical area**. Using high-throughput computational techniques, we discovered that, within a class of biologically plausible hierarchical neural network models, there is a **strong correlation between a model's categorization performance and its ability to predict individual IT neural unit response data**. To pursue this idea, we then identified a high-performing neural network that matches human performance on a range of recognition tasks.*

**Anne:**

- modeling of biologically plausable neural networks: strong correlation between categorization performance and ability to predict individual IT neural unit response data
  - wasn't constrained to match neural data but top layer is highly predicitve of IT spiking responses, intermediate levels = V4 cortex
  - optimization of performance → optimization of IT predictability (but no the other way around)
- Hierarchical Linear-Nonlinear (HLN) hypothesis: higher level neurons (e.g., IT) output a linear weighting of inputs from intermediate-level (e.g., V4) neurons followed by simple additional nonlinearities
- test set: 8 categories with different heavy variations, non-sense scenes
  - humans visual system is robust, but algorithms are not

- Neuroscience study (monkeys) and DNN
    - filters, threshold(like RELU, little more complicated), pooling, normalize → hierarchical stacking
- how good is the model explaining activation of monkeys brain
- HMO → almost perfect in categorization and explain response of IT = human like performance → artificial brain?!
    - top-level predics IT the best
    - middle-level -> V4 best
    - for medium and high variation tasks: existing models break down, Humans and HMO don't!
- DNN model that can recognise objects under strong background, pose & illumination changes rather well (70% correct in high variance condition)
- model's computations are locally build on computations believed to be those of visual cortex
    - capture heterogeneity of visual system: Bypass connections, sub-parts of the network with very different filtering and normalisation parameters, etc.
- DNN as Model for human vision?!

**Quiz: HMO (hierarchical modular optimization) model belongs in the larger class of DNN (Deep Neural Network) models:**

- Essential architectural characteristic = Heterogeneity
    - Many bypass connections and different parameter settings even at the same level of hierarchy
    - Basic operations performed locally are the same throughout network
- Reported a large-scale modelling effort, evaluating around 5000 DNN architectures
- Compared their models both to response of cells in IT Cortex (roughly N = 300, 100 cells) and how well models categorised a set of images (~ N = 6.000 images)
- One critical finding: models optimised for categorisation performance were also superior at explaining variance in IT
- To obtain categorisation performance from HMO Model, linear decode was trained on activity of units at highest levels of the HMO network
    - This showed that the HMO model performed better than computer vision & neruonally inspired models of object recognition on difficult high-variation tasks

*Szegedy et al (2014)* : adversarial attacks → fooling DNNs with small differences of pixels → not recognizable by humans

- shows generalisation errors of DNNs
- carefully designed stimuli → needs knowledge of all weights and connectivities, gradients
- Data augmentation (re-training) → robustness against a specific adversarial attack, but not in general

*Geirhos et al (2018)*: how good are humans and DNNs at recognizing altered images, not engineered stimuli, but weak signals and randomly degraded stimuli (colored, frequency filtered, etc)

- short presentation (200ms), 1/f noise mask (200ms), fast-paced responding(1500ms)
- using several successful DNNs like Res50
- perform superhuman when test and train alteration is the same, but random when not!
- Striking generalisation failure

Prediction is a necessary but not sufficient condition for an idea or theory to be scientific: we need to understand what is going on too. Care is needed when comparing humans to algorithms (or animal species) → similar performance ("behaviour") in one condition ≠ similar performance ("behaviour") in different condition. DNN are bad at generalisation, Comparison to Human overstated

# 11 Scene Perception

## Limits of object recognition

- Visual system cannot recognise multiple objects simultaneously
- For object recognition to be useful for complex scenes we need more mechanisms:
  - Scene perception (gist)
  - Saliency (guiding eyes in a scene)
  - Attention (selection mechanisms more generally)

## Scenes

Gist of scene can be identified very quickly before details are perceived or even object recognized e.g beach scene, street scene,…

➔ Two pathways to scene perception

## Two Pathways

- **selective pathways**
  - allocation to one or few objects at a time and is governed by the attentional bottleneck → selective processing of objects
- **non-selective pathways**
  - processes visual scenes holistically encoding scene gist, spatial layout and ensemble statistics very quickly (no bottleneck by attention)
  - representations are generated as a whole and do not include descriptions of individual objects within the scene
  - has connections with the selective pathway and can for instance guide visual search for particular objects in a scene by helping the observer restrict attention to particular location in the scene

➔ **Spatial Layout**: description of structure of a scene (enclosed, open, …) without reference to specific objects in the scene

➔ **Ensemble Statistics** as explaination for non-selective pathway

- the **average** distribution of properties like **orientation, color over a set of objects / region in scene** → representing **knowledge about** properties of a **group of objects** rahter than individual objects themselves

## Thorpe – Rapid Animal Detection, Feedforward (quiz)

- **human observers** could decide whether a previously unseen **photograph** of a natural scene contained an animal or not
- median reaction time (RT) = **400-500 ms** with mean **90-95%** correct (note **slight** speed-accuracy trade-off)
- **ERP** analyses showed **~150 ms** after stimulus onset the measured neurophysiological correlate could already reliably signal the presence or absence of an animal in a post-hoc analysis → processing completed after quite short time
- This stems from a **feedforward mechanism** theory for object recognition **which argues against** requiring explicit **image segmentation** steps prior to recognition
  - Segmentation assumed to require **time consuming iterative algorithms**

## Background & Motivation

- Many models of contour extraction & object segmentation rely on feedback connections lateral competitive or co-operative interactions however
- There is evidence URAD is in absence of attention (Li, VanRullen, Koch & Perona, 2002)

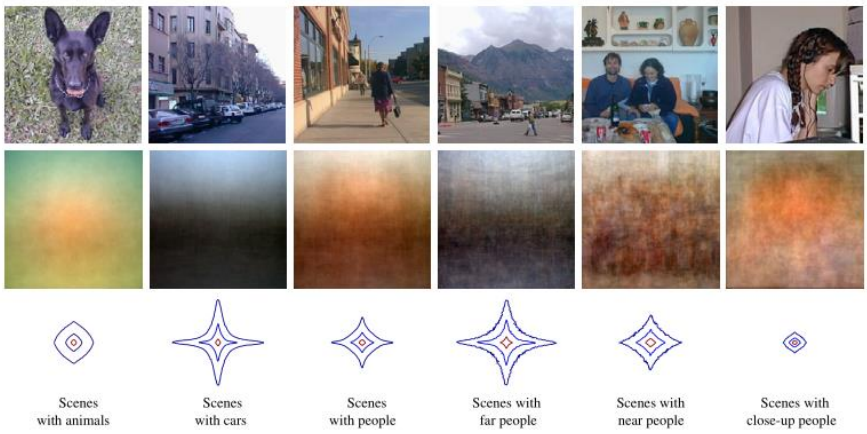
## Statistics of natural image categories

Torralba, Olivia, 2003

- statistical properties of natural images belonging to different categories and their relevance for scene and object categorization tasks
- how second-order **statistics are correlated with image categories, scene scale and objects**
- propose how scene categorization could be computed in a **feedforward manner** in order to provide **top-down and contextual information very early in the visual processing chain**
- resulting: how **visual categorization based directly on low-level features**, without grouping or segmentation stages, can **benefit object localization and identification**.

Quiz:

- Power spectrum of natural (images) scenes is only **isotropic** if averaged across image categories but if analysed separately for different image categories they found **strong correlations between the shape of the power spectrum and image categories**
- Typically a **density plot of the power spectrum of a man-made** image scene is more **star-shaped**
- Based on a **small number of component(s)** of PCA performed on the power spectrum T&O were able to **correctly categorise images** into Yes and No scenes **in 80%** of the cases
- Calculating the PCA of the power spectrum is a **non-linear operation**. Still this operation could be performed in a **feedforward manner**
- Conclusion: Yes vs. No-categorization is so rapid because their **summary statistic** approach does not require an explicit **image segmentation** step




results coincide with Thorpe showing that cognitive tasks (like URAD) being performed in a feedforward way without need of sequential focus of attention or segmentation stages

➔ **natural images can be classified** „Object and Non-Object" at a **success rate of 80% using nothing but global image statistics like power spectrum** → visual system build good template of the features associated with a catergory then using this to make accurate preemptive categorization

**BUT it has not been shown that our visual system actually uses this (spectral differences)**

## Counter-Argument – Wichmann, Drewes, Rosas, Gegenfurtner

- Thorpe 1996: observers can detect animals in images of natural scenes rapidly, but it is still unclear which image features support this rapid detection
- Torralba 2003: a simple image statistic based on the power spectrum allows the absence or presence of objects in natural scenes to be predicted
- Based on psychophysical experiments and computational analyses
- Do observers use the power spectral differences between image categories (animal detection in natural scene)?
    - Performance independent of power spectrum

    - Ease of classification correlates with proposed spectral cue without being caused by it → Hypothesis: animal images are pre-segmented from photographers which causes the power spectral differences and thus aiding URAD
    - High-spatial frequencies appear correlated but not causally related to URAD
- For human observer, animal detection in typical photographs of natural scenes is independent of the power spectrum
- in typical commercial databased the statistics of images may not be as natural due to photographs typically representing a biased view of the world

URAD works with single fixation: the quality of the eye, both optics and sensor and is dramatically non-uniform over the visual field

### Conclusion

- URAD is independent of relative magnitude of high spatial frequencies
- Global image statistics (phase spectra, etc) fail at URAD
- **Ensenmble statistics underly rapid perception of gist of a scene but not (sometimes) very rapid detection within scene**

## 12 Visual Attention

Attention: any of the very large set of selective processes in the brain

- to deal with the huge amount of input (all at once) the NS has evolved
- **mechanisms that are able to restrict processing to a subset of things, places, ideas or moments of time**

**Selective Attention:** form of attention involved when processing is restricted to a subset of the possible stimuli

### Varieties of Attention

- **External**: attending to stimuli in the world
- **Internal**: attending to one line of thought over another or selecting one response over another
- **Overt**: directing a sense organ toward a stimulus like turning your eyes or your head
- **Covert**: attending without moving eyes / head / body
- **Divided**: splitting attention between two different stimuli
    - limited divided attention: cannot read left-hand and right-hand sentences at the same time
- **Sustained**: continuously monitoring some stimulus

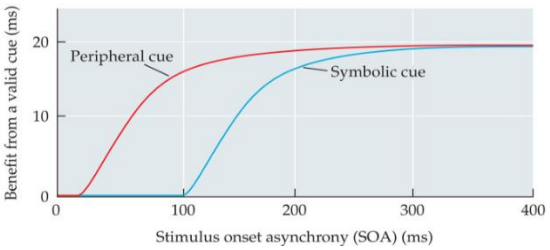
<u>In selective attention:</u>

- **alerting**: maintaining vigilance, readiness to respond, arousal, attention to when a target will occur
- **orienting**: selecting a location or modality to attend
    - brain regions involved in orienting to visual stimuli also involved in orienting to stimuli in other modalities
    - orienting to a location (can) provide better information access from all modalities at that location not only from the expected modality (*Driver, 2004*)
- both are thought to be subserved by separate brain networks

### Selection in Space

- **Classic paradigm**: Posner's cueing paradigm → detect a simple target as quickly as possible
- **Reaction Time** (RT): a measure of the time from the onset of a stimulus to a response
- **Cue**: a stimulus that might indicate where (or what) a subsequent stimulus will be
    - valid, invalid or neutral

- **Exogenous cue**: in directing attention, an exogenous cue is located out (exo) at the desired final destination of attention
    - **Valid cue** [ ] + [ ] -> [ ] + [ ] -> [x] + [ ]
    - **Invalid cue** [ ] + [ ] -> [ ] + [ ] -> [ ] + [x]
- **Endogenous cue**: in directing attention, an endogenous cue is located in (endo) or near the current location of attention
    - **Valid cue** [ ] + [ ] -> [ ] + [ ] -> [x] + [ ] (yellow indicating left)
    - **Invalid cue** [ ] + [ ] -> [ ] + [ ] -> [ ] + [x]
- **Stimulus onset asynchrony (SOA)**: the time between the onset of one stimulus and the onset of another




- **RTs shorter on valid cues**
- **RTs longer on invalid cues**

➔ Effect of cue develops over time
➔ Peripheral cue faster than Symbolic

## Theories of Attention

- **Spotlight Model**: Attention restricted in space, moving from one point to the next, areas within spotlight receive extra processing
- **Zoom Lens Model**: attended region can grow or shrink depending on the size of area to be processed

## Visual Search

Looking for a target in a display containing distracting elements e.g Finding weeds in lawn, remote control on table, …

- **Target**: goal of visual search
- **Distractor**: any stimulus other than the target
- Set size: number of items in visual search display
- Efficiency of visual search: avg. increase in RT for each item added to display
    - Measured by search slope or ms/item
    - The larger search slope or ms/item the less efficient the search
    - Some searches efficient → small slope, some inefficient → large slopes

## Feature Search

Search for target defined by single attribute (like salient color or orientation)

- Quite efficient!
- **Salience**: vividness of a stimulus relative to its neighbours
- **Parallel**: referring to the processing of multiple stimuli at the same time
- Many searches inefficient actually—serial self-terminating search: a search from item to item, ending when item is found
- Familiarity affects performance

## Guided Search

Attention is restricted to a subset of possible items based on information about item's basic features (color or shape)

## Conjunction Search

Search for a target defined by the presence of two or more attributes

- No SINGLE FEATURE defines a target rather defined by co-occurrence of two or more features e.g big, round, red tomatoes

**Scene-based guidance:** In real world searches, real world guides visual search → information in our understanding of scenes that helps us find specific objects in scenes

- A mug will typically be found on a horizontal surface (like a table) and a picture will be found on a vertical surface (like a wall)

**Binding Problem**: challenge of tying different attributes of visual stimuli (which are handled by different brain circuits) to the appropriate object so we perceive a unified object

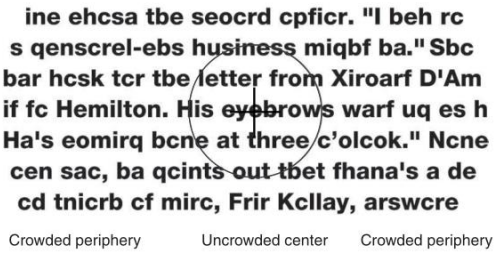
- A vertical red bar moving to the right—how to combine these features when perceiving bar?

## Top-Down – Visual Attention, Feature Binding

- *Rosenholtz et al., 2012*: "**Common Wisdom**" in visual perception one of many **rules of attention is to bind individual features into objects**
- *Treisman&Gelade,1980*: **Feature integration Theory**, is one of the most prominent and most often cited theories in visual perception
    - A **limited set of basic features (color, orientation, size) can be processed preattentively in parallel** → search slope: 0 ms/item & correct target present/absent trials are parallel
    - **other properties**—including the correct binding of features to objects—**require attention**
    - **Preattentive Stage**: processing of stimulus occuring before selective attention is deployed to that stimulus
- Simple targets (don't need attentional binding -> preattentive) can be processed by visual system in parallel
- Complex targets require attentional binding → serial search

## Crowding

Phenomenon where it is more difficult to identify an object when surrounded by flanking objects e.g: DCO + C → is a bottleneck of object recognition




Not masking: detecting the presence of object is not affected (bottleneck of recognition)

Primarily mediated by cortical neurons: flankers in a different eye to the target are almost as effective at causing crowding as flankers in the same eye

Primarily perpheral but can be found in fovea too (really small objects)

Crowding zones scale with eccentricity according to "Bouma's Law":
flankers begin to cause crowding at approximately half the retinal eccentricity of target
Rosenholtz: peripheral vision as compulsory texture perception

## What is wrong with Feature Integration Theory?

- Ignores the profound inhomogenity of the visual system: strikingly different resolution across the visual field → human visual sensitivity is not uniform (fovea, peripheral)
- *Rosenholtz 2012, Geisler&Chou 1995*: serial vs. parallel search is simply a function of how easily the target & distractors can be discriminated in periphery
    - Serial search if needing to foveate on a target / distractor to distinguish
    - Parallel search if difference is robust enough to visible with low resolution periphery

# 13 Visual Saliency

## Saliency

*Itti, 2007:* Distinct subjective perceptual quality which makes some items in the world stand out from their neighbours and immediately grab our attention

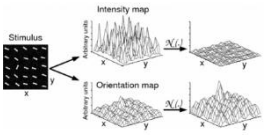
- Important for complex biological systems to rapidly detect potential prey, predator or mates in cluttered visual world
- Complexity of processing all stimuli in the visual field simultanously is prohibitive
    - Solution: restrict complex object recognition process to small area / a few objects at any one time
    - Solution: Serialisation of visual scene analysis after gist is understood quickly
        - Problem: by processing one region / object at a time -> how to select next target of processing?
- Early vision → distinct subjective perceptual quality which makes some stimuli stand out from others → brain evolved to compute salience rapidly and in an automatic way (in real-time over entire visual field)
- CORE: bottom-up stimulus driven signal, announcing "this location is sufficiently different from its surroundings to be worth of attention!"
- Sometimes (carelessly) described as physical property of a visual stimulus—instead it's the consequence of an interaction of a stimulus with other stimuli (and visual system)
    - e.g color-blind person has very different experience of visual salience than a person with standard color vision even when looking at the exact same scene

➔ saliency computed automatically, effortlessly and in real-time

➔ highly-salient things automatically draw attention

## Essence of Salience

### Competing for representation

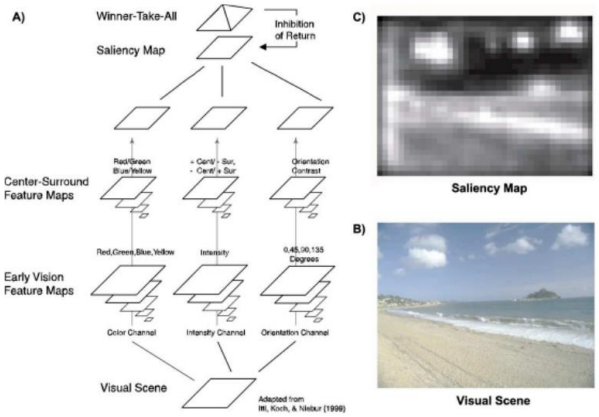



Differences to each other, Intensity same, Orientation not the same → attention drawn

### First (well-known) Visual Saliency Model – Itti & Koch Model

*Itti&Koch, 1999: a saliency-based search mechanism for overt and covert shifts of visual attention*

- Most models of visual search are based on **concept of saliency map**: **an explicit 2D-map that encodes saliency or conspicuity of objects in visual environment**
- **Competition among neurons** in this map leads to a **single winning location** which corresponds to next attended target
- Inhibiting this location, automatically allows system to attend next most salient location

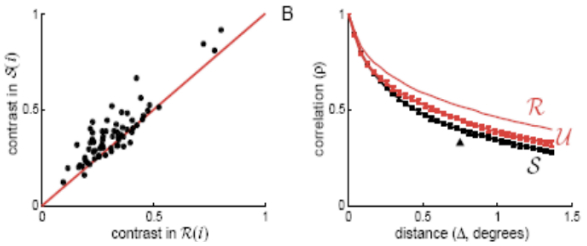



Model describes this in computer implementation → Focusing on: **Problem of combining information across modalities** (here orientation, intensity, color information), purely stimulus-driven

Model applied to common psychophysical stimuli and very demanding search task → performance addresses how primate visual system carries out visual search (with one ore more saliency maps)

What is special about local image statistics at fixation points? ( whats p(fixation|image) )

## Statistical Properties of Fixation Locations




*Reinagel&Zador, 1998:*

Center Pixel more different to surrounding pixels in fixation patches

correlation coeff. of RMS and model output = 0.69

*Krieger, 2000:* 3rd-order statistics -> energy distribution is more circular: the saccadic **selection system avoids** image regions which are dominated by **single oriented structure, while selecting different orientations** (occlusions, corners, etc.) → Sky and Clouds example

### Machine Learning

Algorithmic approach to the science of learning from data, initially mainly developed in computer science, statistics is the science of learning from data too → both fields are identical in intent

- At its best, excels at discovering hidden structure in existing data in order to predict novel data (prediction)
- At its worst, designs algorithms nobdy wants to use

**Algorithms**

- CS Good: low complexity (fast), low space complexity (small memory),…
- Statistics Good: consistency → with increasing input data, algorithm converge to true solution, confidence → algorithm should know about its own reliability
- CS vs Statistics don't necessarily align => find a good trade-off

### ML Approach to Visual Saliency

Previously: top-down, machanistic modeling approach develeoping biologically inspired models using neurophysiological hardware like gabor-filters this was well-suited and time-proven IF abundant domain knowledge were available—not often the case in sensory psychology! Ad-hoc choices have to be made (exact filter types, sizes, combos…)

→ ML Approach is **to construct model from data:**

- very general model class that does not know about the problem but can adapt well to large class of problems
- numerically learn (optimise) ist parameters so new data is predicted best

### Data Representation

For each fixation location (data point, i = 1…36.000) store local pixel values in feature vector $X_i$ and associate a label $Y_i = 1 / -1$

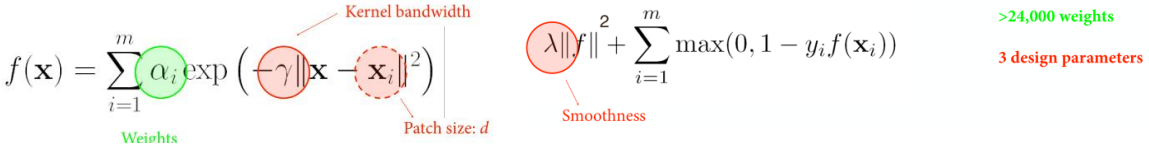
**„Non-Fixed" Patch:** Generate Background Examples with same spatial distribution as fixations

### ML Method

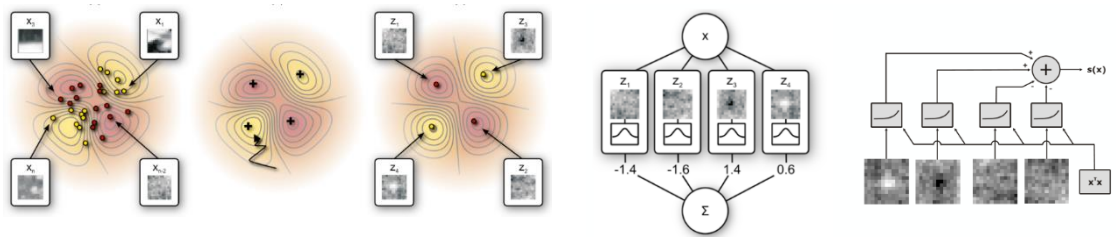
Make model class as general as possible:

- model is radial basis function (RBF) network with one basis function centred on each training example („Non-parametric" as ist complexity grows with number of Data Points)
- General -> universal approximation property, no preference for any image structure, no knowledge about shape or size of receptive field in HVS
- Compute weights a using hinge loss + L2-Regularizer (=SVM): find a is convex = efficent and guaranteed to find global optimum
- Find Design Parameters Lambda, Gamma and Patch Size d via exhaustive grid search using cross validation estimates of accuracy (feasible only in 3D)

**Radial-Basis-Function Support Vector Machine (RBF-SVM)**




**RBF-SVM after optimization**




a lot of parameters to keep in mind ~13x13 Dim. → „Height Map“ or „Mountain Map“—high saliency = Peaks +, low saliency = Trough - → Student found way to reduce to 4 Key Parameters/Weights leading to 2 Peaks and 2 Troughs: these were resembling On-center / Off-Center Cell Behaviour (Lateral Inhibition)

**Generalization to Novel Data Set:**

| s.e.m | Ground Truth | Novel Data |
|---|---|---|
| ML-Model | 0.64 +- 0.010 | 0.62 +- 0.012 |
| Itti-Koch | 0.62+- 0.020 | 0.57+- 0.020 |

## Interim Conclusions

- Bottom-Up Saliency inferred from data without prior assumptions…
- Most relevant regularity in local image structure at fixation is a simple center-surround configuration (biologically plausible but learned only from data and not assumed!)
- Assembly: Small Network with only 4 Linear Receptive Fields followd by static non-linearity and contrast gain-control → Prediction Performance of Full RBF-SVM—this model is relatively simple compared to Itti & Koch
- This: extended psychophysical receptive or perceptive field analysis recovering perceptive field / decision image networks
  - ⇨ System Identification via reverse-engineering a non-linear Kernel Machine

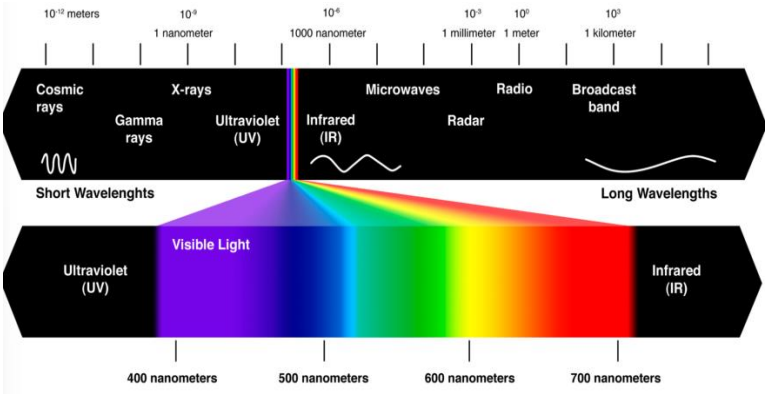
## Summary

- Saliency and Modelling it has garnered a lot of interest
  - technological, commercial application: navigation, autonomous driving, military, video compression, advertising
- difficult Problem because human fixation patterns are clearly stochastic but almost all current models are deterministic (see Itti&Koch and Spatial Point Process)
- current successful models are Deep Convolutional Neural Networks, using high-level features such as faces (Question arise if they really measure Bottom-Up-Saliency)
- SVM-Based ML approach suggests Bottom-Up Component in images may be based on rather simple contrast-normalised centre-surround computation (Kienzle, 200)

# 14 Colour Vision 1 – Psych, Light and Reflectance Funct., Principles

## Perception

Colour not a physical property, purely psychological




Eye has tree types of cones which convert light into neural signal → transformed into opponent colours in Retinal Ganglion Cells → in brain these excitation patterns are interpreted as colours.

Colour is the Sensation allowing us to distinguish two surfaces of the same brightness, without any structural information

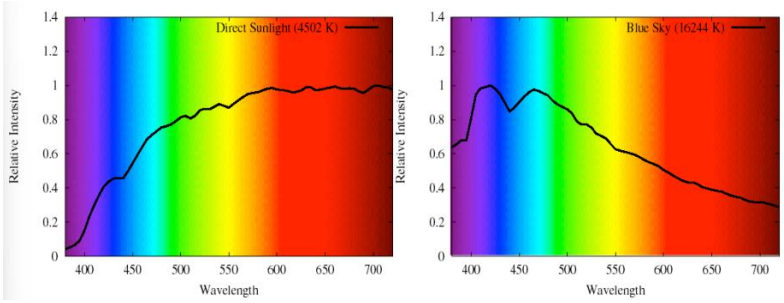
At least 2 mil. Colour shades distinguishable: 200+ shades -> hue, 20+ saturation levels, 500+ brightness values

**Superposition of Light:** Spectral Power Distributions obey Superposition (just waves)

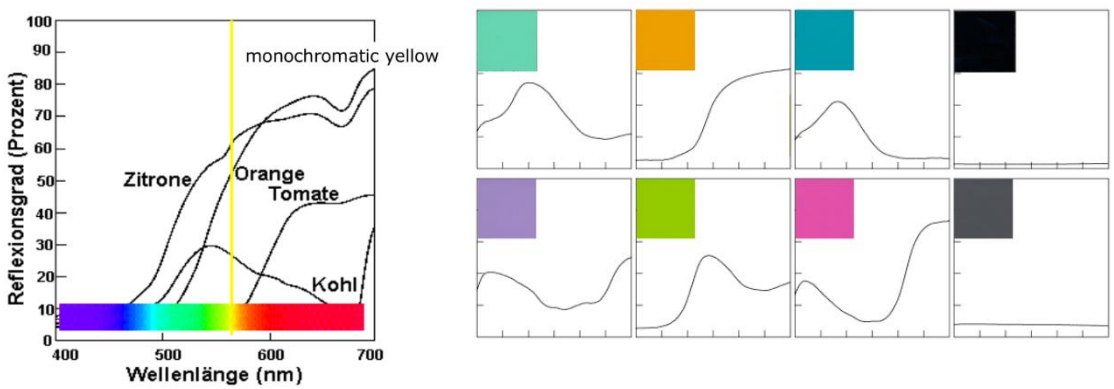
**Reflectance and Absorption**

- Part of incident Light is reflected, the other absorbed
- spectral composition of light (which is caught by the eye) changes when illumination changes
- **Spectral Power Distribution:** distribution over energy over wavelength, property of light
- **Spectral Reflectance Function:** Proportion of Energy at each wavelength reflected by surface, property of surface

**Natural Spectra**




**Spectral Reflectance Functions**



## Basic Principles

- **Additive Color Mixing**: a Mix of Light, adds coloured lights
  - Light A, Light B → both are reflected from surface to eye → effects of those two lights add together e.g Light A = Blue, Light B = Yellow -> A + B = White = Additive Color Mix
- **Subtractive Color Mixing**: a Mix of Pigements, takes light away
  - Pigment A, Pigment B → light shining on the surface will be subtracted by A and some by B → Remainder contributes to Color Perception e.g White Light -> yellow filter on W = yellowish - blue filter on W = blueish == green remainder
- Most light we see is reflected, illumination is provided by sun, light bulbs, fire…

## Three Steps To Colour Perception

1. **Detection**: Wavelengths of light must be detected
2. **Discrimination**: we must be able to tell the difference between one wavelength and another
3. **Appearance**: we want to assign perceived colours to lights & surfaces in the world, those perceived colours be stable over time regardless of different lighting conditions

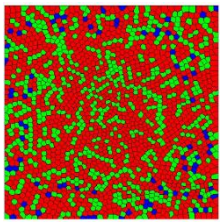
### Step 1: Colour Detection

**Light Detection by Photoreceptors:**

| | | | |
|---|---|---|---|
| S-Cones | Short Wavelengths | "blue" | 420 nm |
| Rod | | "Dark" | 498 nm |
| M-Cones | Medium Wavelengths | "green" | 535 nm |
| L-Cones | Long Wavelengths | "red" | 565 nm |

L-Cones Peak is 565nm corresponding to yellow not red!

**Cone Mosaic:**



Three Types of Cones in human retina which are arranged in a mosaic:

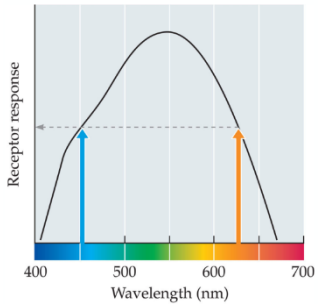
- at one location in space only one type of cone, for most:
- roughly twice as many L than M cones
- only abt. 10% of all cones are S Cones and none in the central fovea

**Detection at Different Illumination Levels**

**Photopic**: Light Intensities that are bright enough to stimulate the cone receptors and to saturate rod receptors to their max. response → Sunlight and bright Indoor Lighting are photopic

**Scotopic**: Light Intensities that are bright enough to stimulate rod receptor but too dim for cone receptors → Moonlight, extremely dim Indoor Lighting are scotopic

### Step 2: Colour Discrimination




Single Photoreceptors show different responses to lights of different wavelengths but same intensity

Principle of Univariance: Light Absorption via opsins → single opsin type maps all visible light to 1D Response, this perfectly confounds lights wavelength with its intensity

Light at lower WL (blueish) has more energy per quanta than High WL (redish), the opsin reaction is the same for absorption of each quantum

➔ An infinite set of different wavelength-intensity combinations can elicit exactly the same response from a single type of photoreceptor --> one type of Photoreceptor cannot make colour discriminations based on WL

## Scotopic Condition - Colourless At Night

- Rods sensitive to scotopic light levels
- All rods contain the same photopigment molecule: rhodopsin
- All rods have the same sensitivity to various WL of light
- Rods obey principe of univariance and cannot sense differences in color, when scotopic: only rods active → so world seems colorless

## Photopic Condition – Colour Discrimination

Trichromacy: Theory that color of any light is defined in our visual system by relationships of three numbers = three cone outputs → in photopic: we avoid principle of univariance → changing intensity of light, changes the absolute cone activiations but not their relative relationships

- Discovered by Young-Hemlmholtz via psychophysical observations then in 1990s with anatomical base

## Color Matching

*James Maxwell, 1831-1879:* developed color-matching technique, still used today
**Modern Version**: two sides of bipartite field can be made indistinguishable for every test light using three primary lights. If the three primary lights are independent -> two are not enough, more than three is not needed → Trichromacy: Human Colour Perception based on three independent mechanisms (3D)

## Metamerism

Test light can have any spectral power distribution but mixture of primaries can only contain spectra given by weighted sums of three primary lights—yet human can set these lights to appear the same → **Metamers**: physically different stimuli appearing the same, here: different mixes of WL that look identical e.g Red + Green = Yellow but physically different than just Yellow!

## Trichromacy

Behavioural Concept: color space in color matching experiment is 3D, behaves like linear vector space (does not mean species has three photoreceptors) e.g digital cameras only have single photo sensor but 3+ colour filters and birds use coloured oil

## Colour as linear 3D Vector

*Grassman Law, 1853*: Color Matching is linear -> IF R1(.15) + G1(.25) + B1(.5) = Cyan & R2(.5) + G2(.15) + B2(.5) = Magenta THEN (R1+R2) + (G1+G2) + (B1+B2) = Cyan + Magenta

- The three primaries lin. Independent, form a basis in 3D then any point in space can reached (with negative coeff.)—if not then shine one (or two) primaries to the target to make the two halves indistinguishable

## Colour Spaces

(in Trichromatic Theory) All perceivable colours can be created by mixing three primaries, colours lie in 3D vector space → many definable colour spaces describing perceivable colours

- **RGB**: defined by outputs of long, medium, short WL lights
- **HSB**: defined by hue, saturation, brightness
- **Hue**: chromatic (color) aspect of light
- **Saturation**: chromatic strength of hue
- **Brightness**: distance from black in color space

**Colour Contrast of Natural Objects**

- High correlation between L and M Cone Signals → not very efficient
- Re-coding: L+M (brightness), L-M (red-green opponent colours)

## Opponent colours

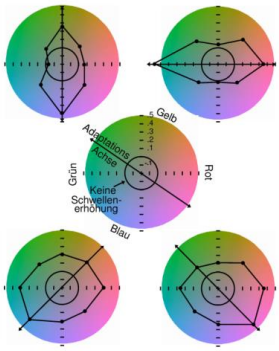
**Chromatic Signal Decorrelation via cone-opponent mechanism**

- Cone Photoreceptors → Retinal Ganglion Cells, LGN cells: L+M = Luminance, L-M = Red-Green, S-(L+M) = Blue-Yellow

**Physiologically-plausible colour space**

- (D)errington-(K)rauskopf-(L)ennie space: Coord. Represent purported responses of three 2nd-site colour discrimination mechanism L+M, L-M, S-(L+M)
- cardinal directions: Modulation directions changing response of one of these while leaving the response of the other two fixed
- Oppenency in Centre-Surround Organisations found in Retina, LGN, Ganglion

**Adaption**

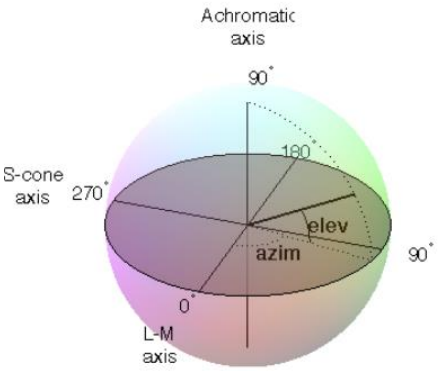



Adaption along one of the opposing color axes leads to reduction in detectability of signals along this axis while signals along the other opposing axes are not

Opponent channels are independent of each other also referred to as cardinal colour directions

# 15 Colour Vision II

Cone-Oppnent Colour – DKL Space

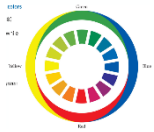



## Opponent Colours: Independence

*Hering*: color combinations -> some „legal“, some „ilegal“

Bluish Green (Cyan), Reddish Yellow (Orange), Bluish Red (Purple) → cannot have Reddish Green, Bluish Yellow

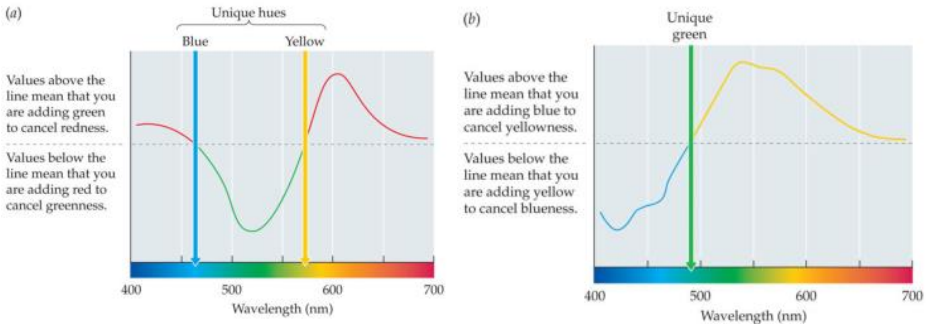
## Opponent Colour Theory



Perception of color depends on output of three mechanisms. Each of them based on an oppenency between two colors:
R-G, B-Y, B-W

→ Quantification? Hue Cancellation Experiments (Hurvich, Jameson, 1957) to measure subjective color appearance (different from discrimination thresholds)

## Hue Cancellation Experiment




Bluish Green + Reddish = Blue (red cancels green out), more green then more red is needed

**Unique hues**: single colour term (B,Y,G,R) e.g unique blue has no hint of red or green
**Red doesn’t have spectral locus**: all L-WL appear red, doesn’t exist as monochromatic light

## Wavelengths of Unique Hues

- Consistent identification of unique hues over several subjects, individually
  → range of 2 nm to 10 nm (Webster, 2000)
- Interpersonally big variation in location of unique hues especially for green
  - *Schefrin, Werner, 1990:* for normal subjects found it between 488 – 546 nm -> range of 49 nm or around 16 % of visible Spectrum

Colour Opponent axes vs cone opponent axes: Cardinal & unique axes only partially agree!

## Individual Differences in Colour Perception

**Qualia**: Private conscious experiences of sensation and perception -> "Is my blue the same as your blue?" General agreement on colors, very large agreement on color matching (color the same?)
**Basic color terms**: single words describing colors and have meanings which are agreed upon by speakers of a language
**Cultural Relativsm**: in sensation & perception the idea of basic perceptual experiences determined partly by cultural environment (various cultures describe color differently)
**Color Vision Deficiencies**: 8% male, 0.5 % female population have some sort of "color blindness"

- **Color-anomalous**: term for "color blindness", most can still make discriminations based on WL, these are just different from the norm
- **Types:**
  - **Deuteranope:** absence of M-Cones (Greenish), M:1.1, F:0.01
  - **Protanope:** absence of L-Cones (Reddish), M:1.0, F:0.02
  - **Tritanope**: absence of S-Cones (Bluish), M:0.002, F:0,001
  - **Cone monochromat**: only one cone type → truly color-blind
  - **Rod monochromat**: has no cones of any type → truly color-blind, very visually impaired in bright light, M:0.003, F:0.002
  - **Achromatopsia**: inability to see color due to cortical damage
  - **Anomia**: inability to name objects/colors despite seeing and recognizing them

## Color of Lights to World of Color

A lot of color in scene usually present → Influencing each other

- **Color Contrast**: perception effect where color of one region induces the opponent color in a neighbouring region
- **Color Assimilation**: perception effect in which two colors bleed into eachother, each taking on some of the chromatic quality of the other
- **Unrelated color**: color that can be experienced in isolation
- **Related colo**r: color like brown or gray, which is seen only in relation to other colors e.g gray patch in complete darkness appears white
- **Negative afterimage**: afterimage whose polarity is the opposite of original stimulus
- **Light stimuli**: visual image seen after a stimulus has been removed
- **Colors are Complementary** e.g red produces green afterimage, blue produces yellow afterimages …
- **Color Constancy:** tendency of a surface to appear the same color under fairly wide range of illuminants. To achieve color constancy: discount illuminant and determine what true color of a surface is, regardless of how it appears → **Illuminant**: Light that illuminates surface
  - Surface Reflection * Illuminant = relative light * cone sensitivity = response → how to get actual color of surface?
  - Intelligent guesses about illuminant, assumptions abt. Light sources, surfaces,… visual system seem to use many mechanisms: normalisations: receptor outputs, patches of scene, entire visual field…
  - Visual system "knows": brightness changes across shadow unlike hue (recognition of shadows)

## Reasons of Colour Vision

- Easier identification of food (berries and ripeness, flowers and UVL -> bees)
- Sex: flower color signals bees “nectar” when flowers are ready to bloom (pollination), colorful patterns signal healthy mate (procreation)

## Blue Dress Phenomenon

Perception of dress seem tied to individuals interpretation of illuminant:

- Assumption: in shadow or natural light → dress gold white
    - Because shadows overrepresent blue light thus subtracting S-WL will make image look yellowish similarily: natural daylight also overrepresents S-WL

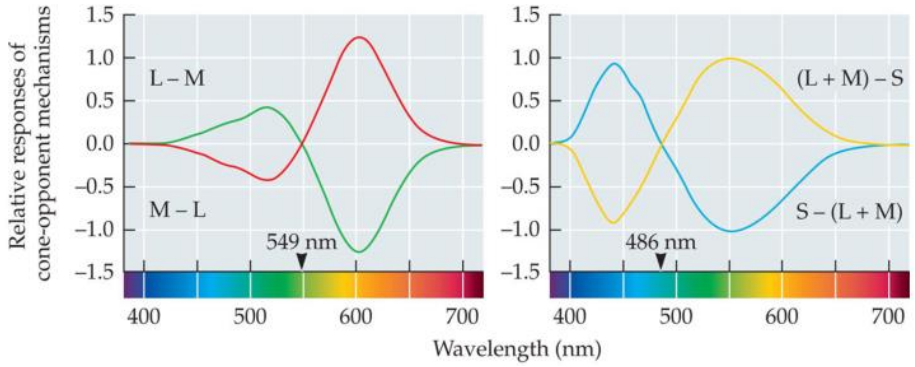
## Summary

- First Stage well described by Trichromatic Theory: 3 Primary Colours -> cone sensitivity => color matching experiments, metamers, color anomalies
- Second Stage is cardinal axes of color space DKL -> cone opponency => aftereffects, color adaption, masking
- Third Stage is color appearance similar to cardinal axes => Hering’s opponent colors, unique hues from hue cancellation experiments
- Additional higher-order effects not explicable

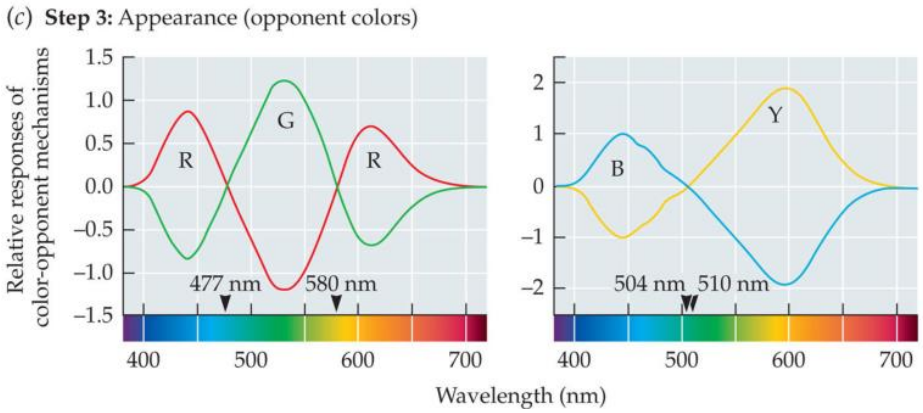
**Step 1: Detection**—S,M,L cones detect Light

**Step 2: Discrimination**—cone-opponent mechanisms discriminate wavelengths

- L-M and M-L compute R vs. G, (L+M)-S and S-(L+M) compute B vs. Y



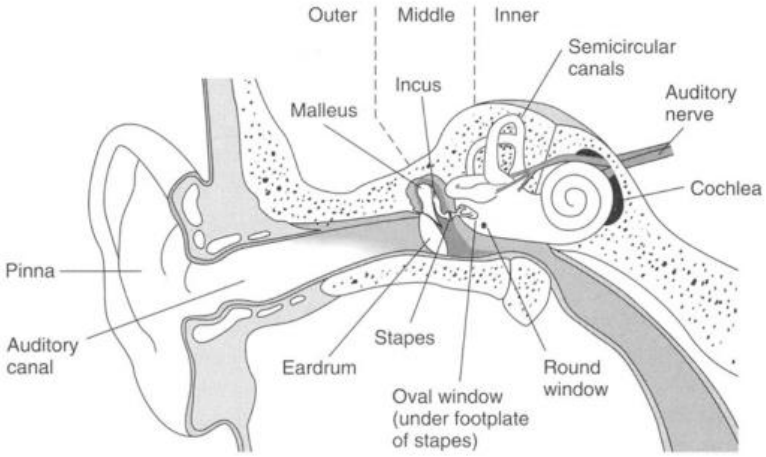

**Step 3: Appearance**—further recombine signals to create final color-opponent appearance

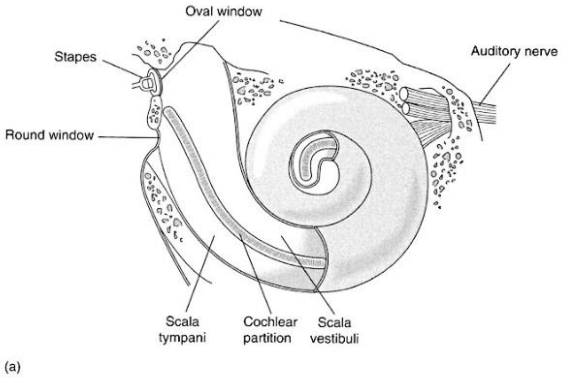


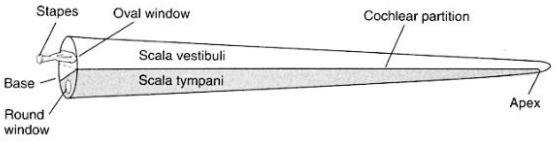
# 16 Auditory System I – Anatomy, Mechanic

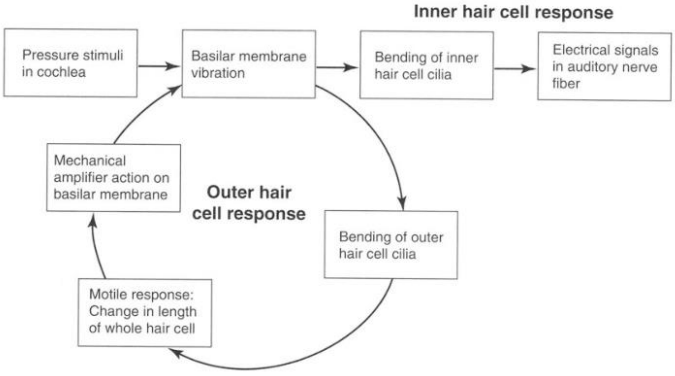
Oscillating, moving, vibrating object -> Pressure Wave in Air -> Vibrations in Eardrum -> Transduction from mechanical osc. To electrophysiological signal -> analysis of those in midbrain, cortex

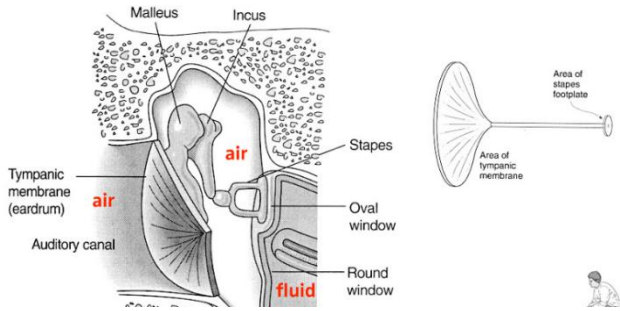
- **Atmospheric pressure**: $10^5$ Pa
- **Pressure change in hearing**: $2 \cdot 10^{-5}$ ... 20 Pa (0 … 120 dB)
- **Hearing frequencies**: 20 ... 20'000 Hz
- **Corresponding wavelength** λ: ~ 20 m ... 2 cm (1'000 Hz ~ 34 cm)
- **Speed of sound** $c=\lambda \cdot f$: 340 m/s
- **length of ear canal**: 30 mm
- **diameter of cochlea**: 10 mm with 2.75 turns
- **length of uncoiled cochlea**: 35 mm (mice: 7 mm, elephant 60 mm), **width of cochlea**: 2 mm
- **width of the basilar membrane**: ~0.1 mm at the base (taut), ~0.5 mm at apex (slack)
- ~3500 inner hair cells & ~12000 outer hair cells / ear
- ~ 30'000 auditory nerve fibres / ear
- ~10 nerve fibers / inner hair cell
- ~ ¼ nerve fibers / outer hair cell
- ~ 90% nerve fibers from inner hair cells, 10% from out hair cells
- maximum spike rate of hair cells ~0.5-1 kHz, phase locking up to 2-4 kHz

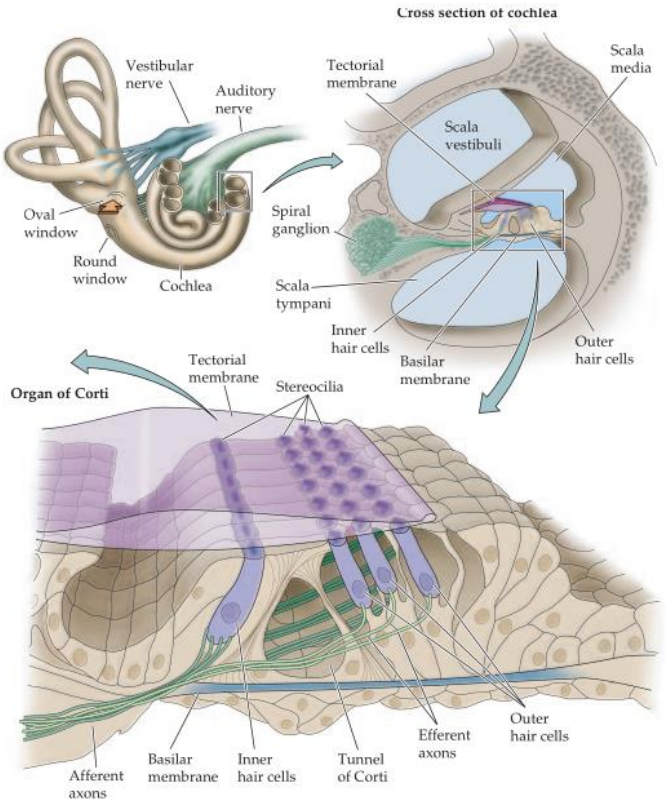


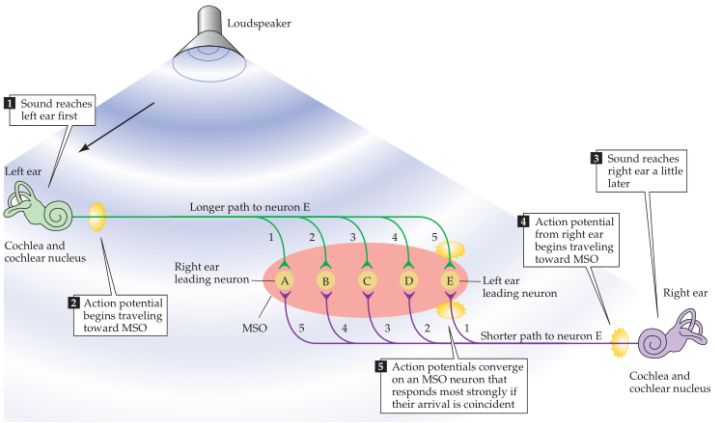



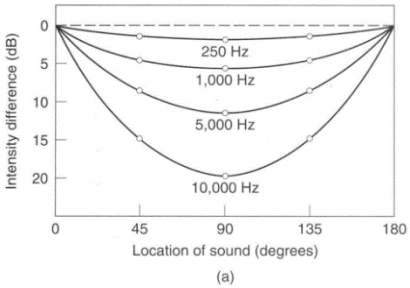



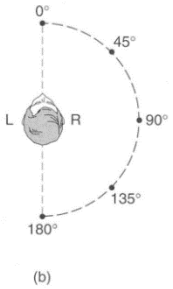



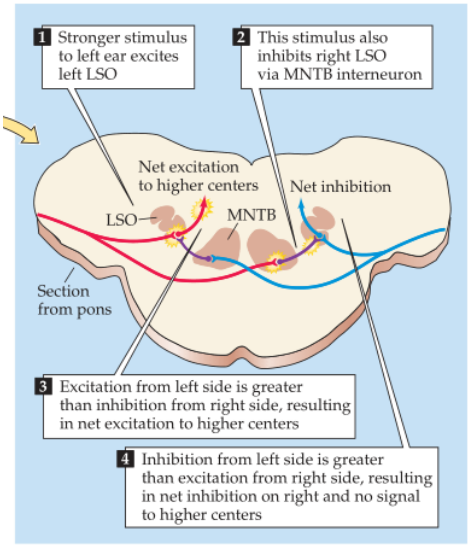
# 17 Auditory System II – Localisation

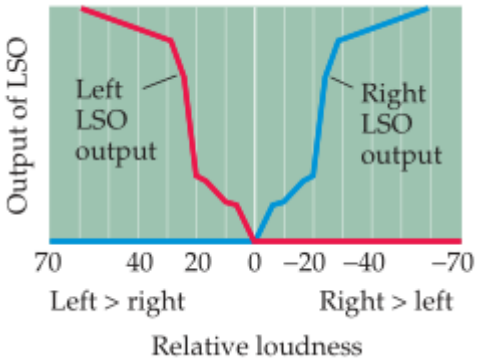
- **Interaural Time Difference**: the time interval between when a sound enters one ear and when it enters the other ear, Jeffres (Goldstein)




- The outer and middle ear transduce and amplify sound pressure waves.
- In the cochlea these pressure waves are transformed into auditory nerve action potentials.
- Highly specialised inner hair cells are the sensory cells for audition, sitting in the organ of Corti.
- Outer hair cells are part of an active amplification mechanism, that also sharpens tuning.
- The auditory system (initially) follows a strictly tonotopic organization
- Sound localization much improved using more than one ear. It relies on time and level
- differences, as well as changes in sound spectrum due to torso/head/pinnae interaction.
- Noise outside the critical band is irrelevant for tone detection.
- Two added tones in different CBs perceived louder than tones in the same CB.
- Tones in different CB don't produce difference tones or beats.
- An analysis into harmonics only works for separations larger than CB.
- Binaural signal comparison only happens within one CB.
- Relative phase of tones in different CBs perceptually irrelevant.