Models of early visual perception

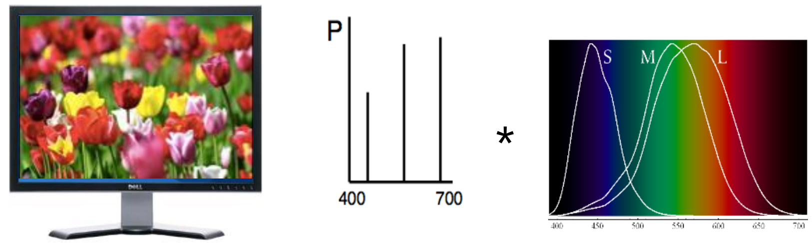
Advanced Graphics and Image Processing

Rafal Mantiuk

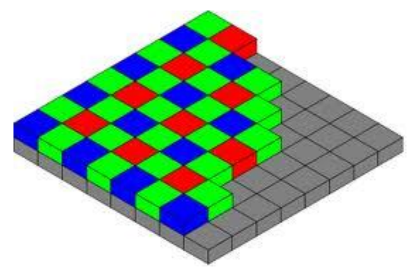
Computer Laboratory, University of Cambridge

Many graphics/display solutions are motivated by visual perception

Image & video compression

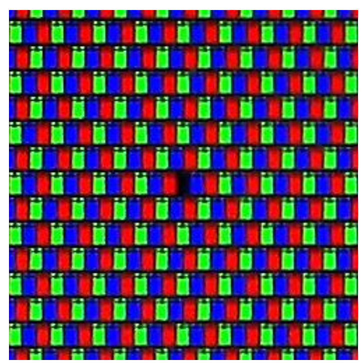


Display spectral emission - metamerism

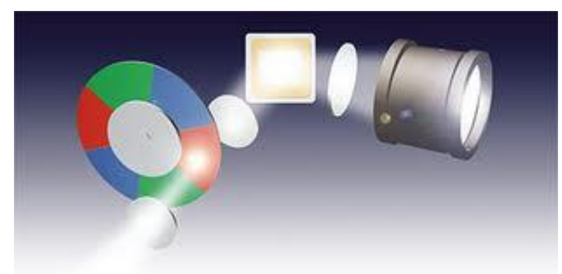


Camera's Bayer pattern

Halftonning



Display's subpixels



Color wheel in DLPs

…

Perceived brightness of light

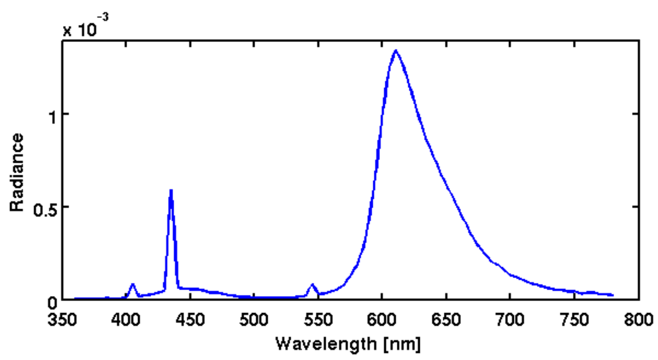
Luminance (again)

- Luminance – measure of light weighted by the response of the achromatic mechanism. Units: cd/m^2

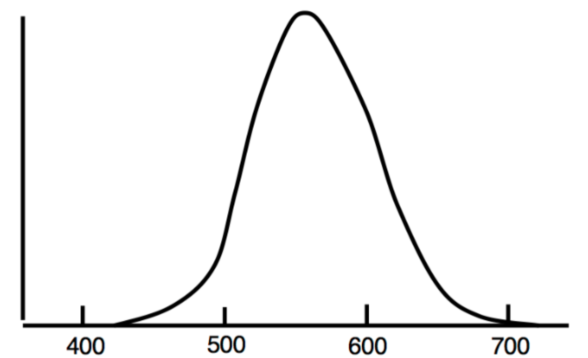
Luminance

$$L_V = \int_{350}^{700} kL(\lambda)V(\lambda)d\lambda \qquad k = \frac{1}{683.002}$$

Light spectrum (radiance)



Luminous efficiency function (weighting)



Steven's power law for brightness

- Stevens (1906-1973) measured the perceived magnitude of physical stimuli
  - Loudness of sound, tastes, small, warmth, electric shock and brightness
  - Using the magnitude estimation methods
    - Ask to rate loudness on a scale with a known reference
- All measured stimuli followed the power law:

$$j(I) = kI^{a}$$

Perceived magnitude

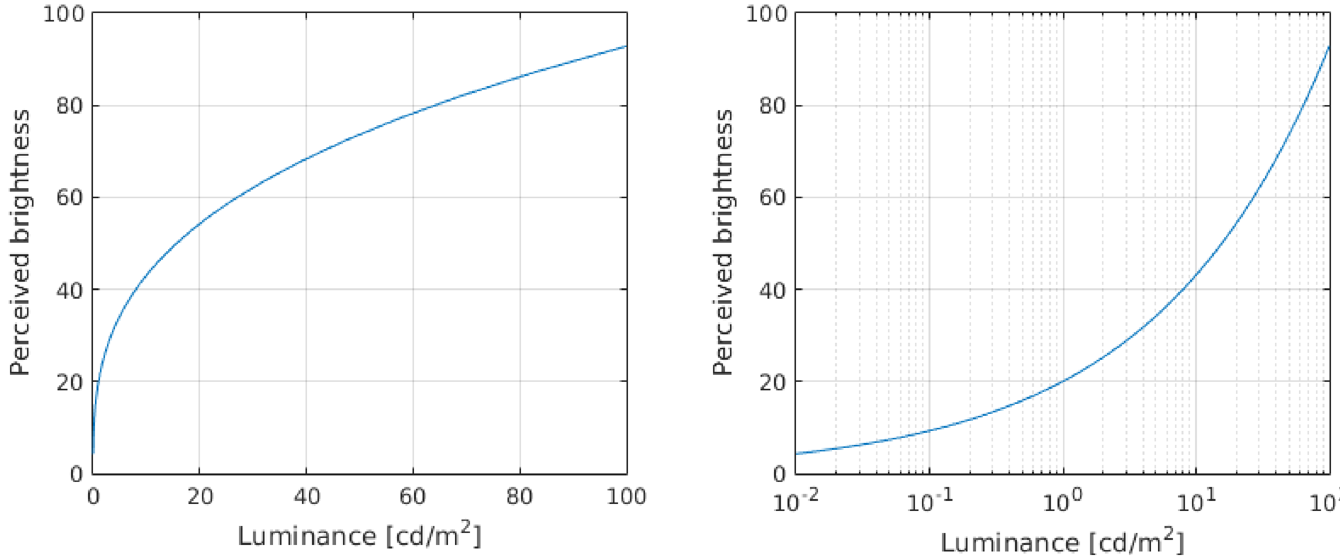
Exponent

Constant

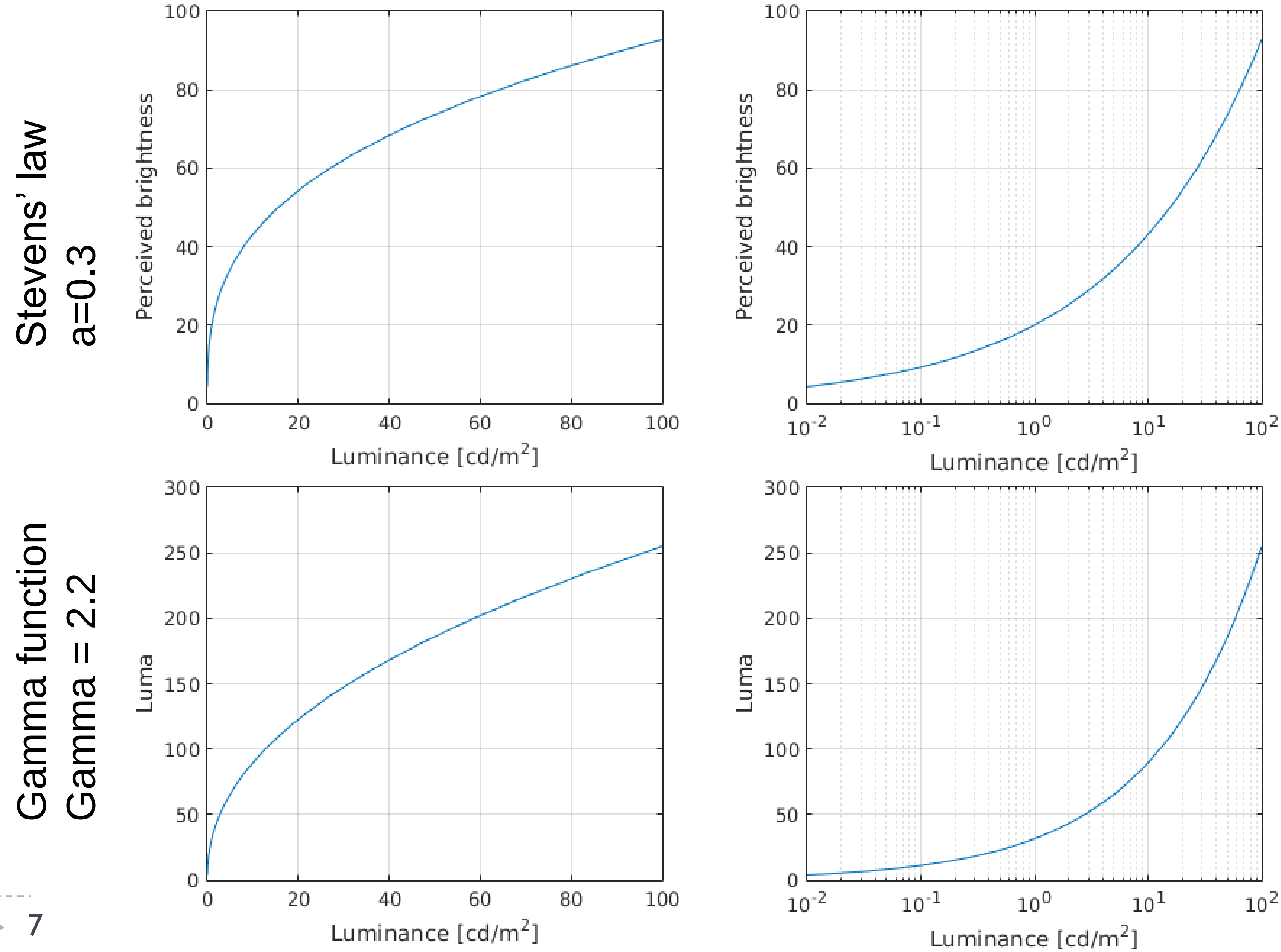
Physical stimulus

- For brightness (5 deg target in dark), a = 0.3

Steven’s law for brightness

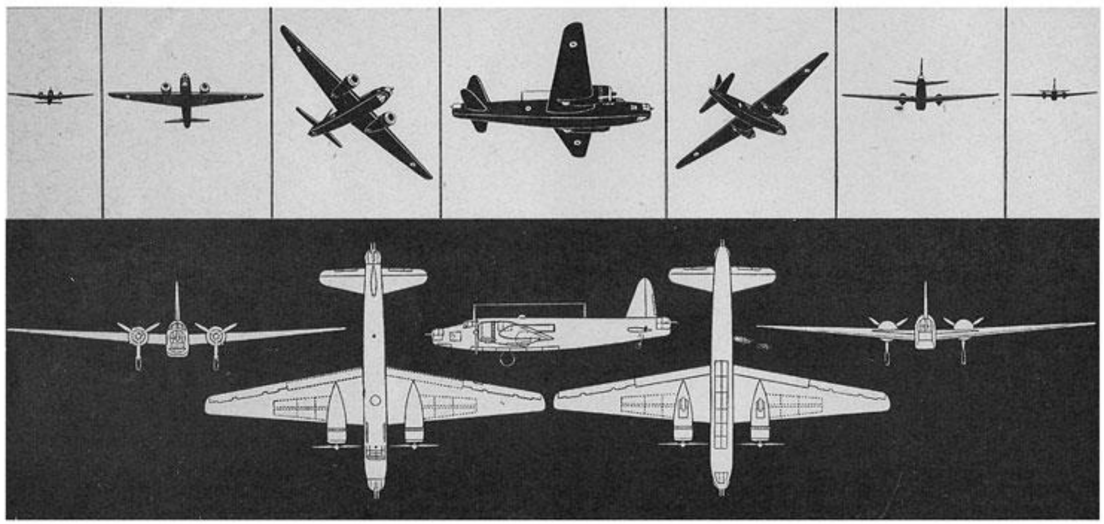


Steven’s law vs. Gamma correction



Detection and discrimination

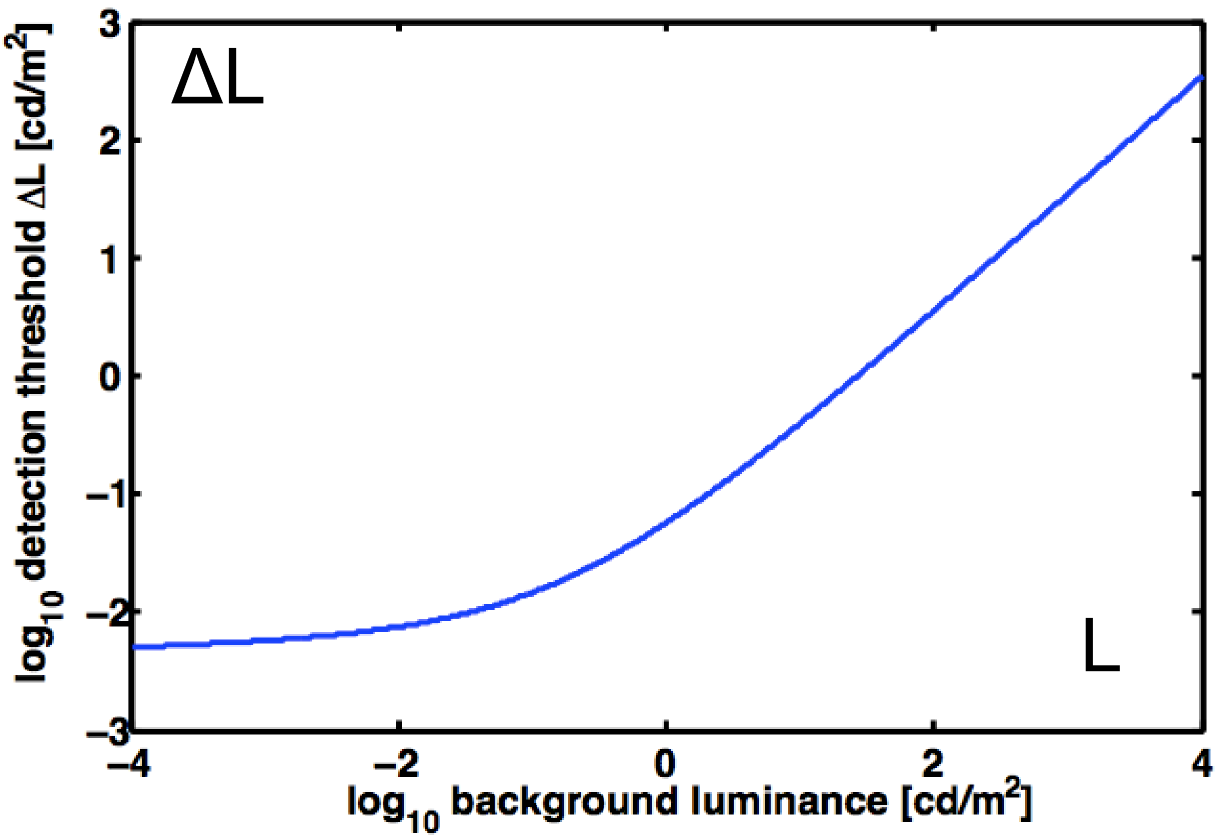
Detection thresholds

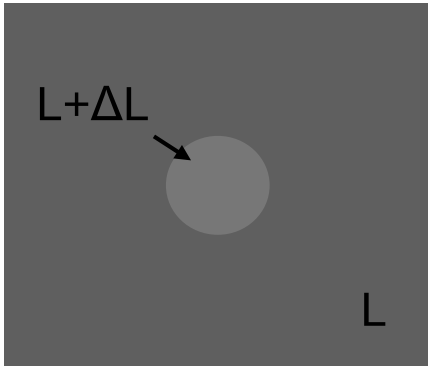


- The smallest detectable difference between
  - the luminance of the object and
  - the luminance of the background

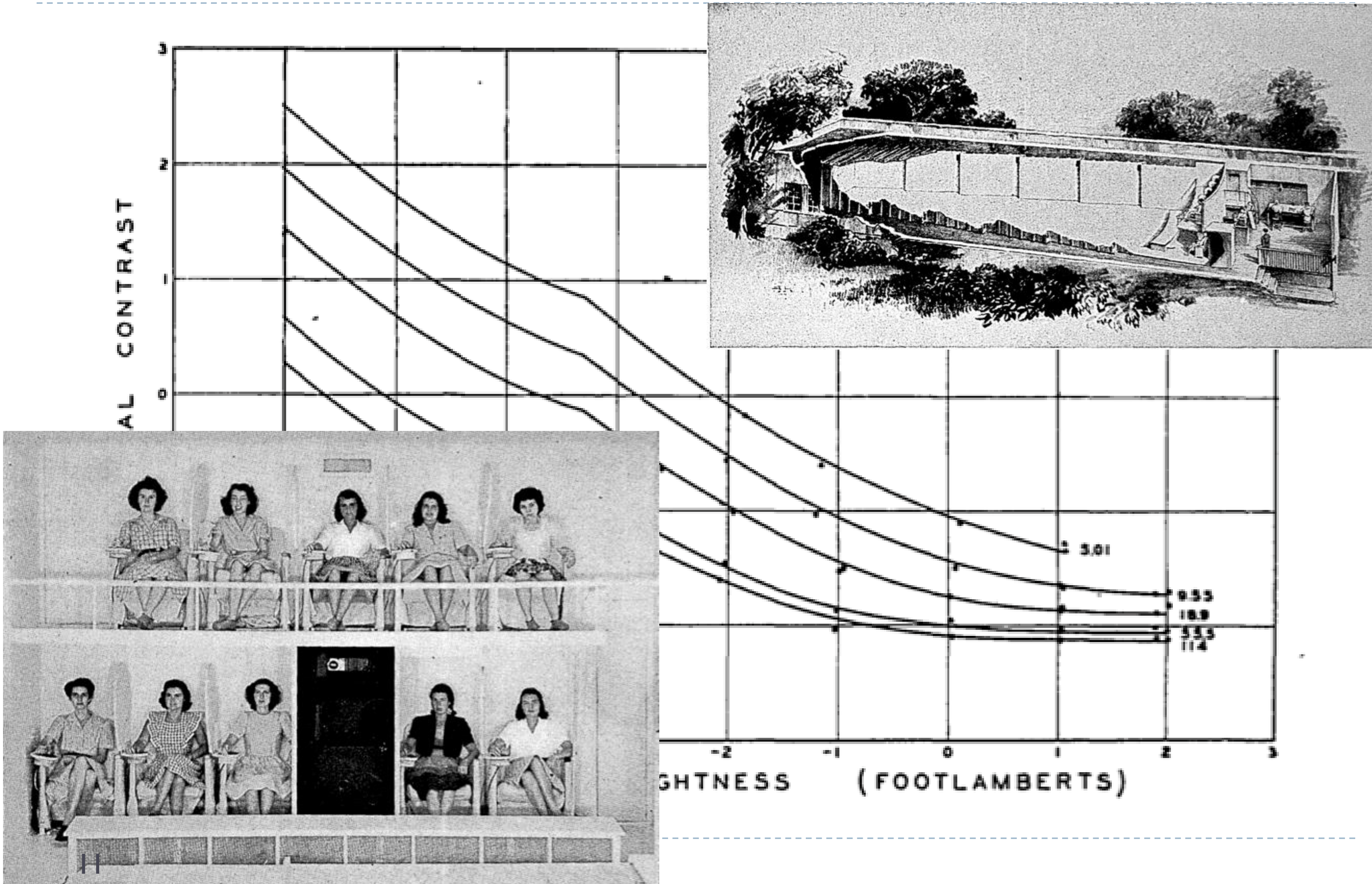
Threshold versus intensity (t.v.i.) function

- The smallest detectable difference in luminance for a given background luminance

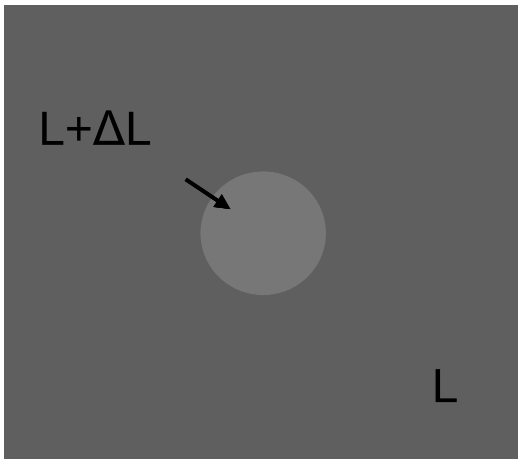


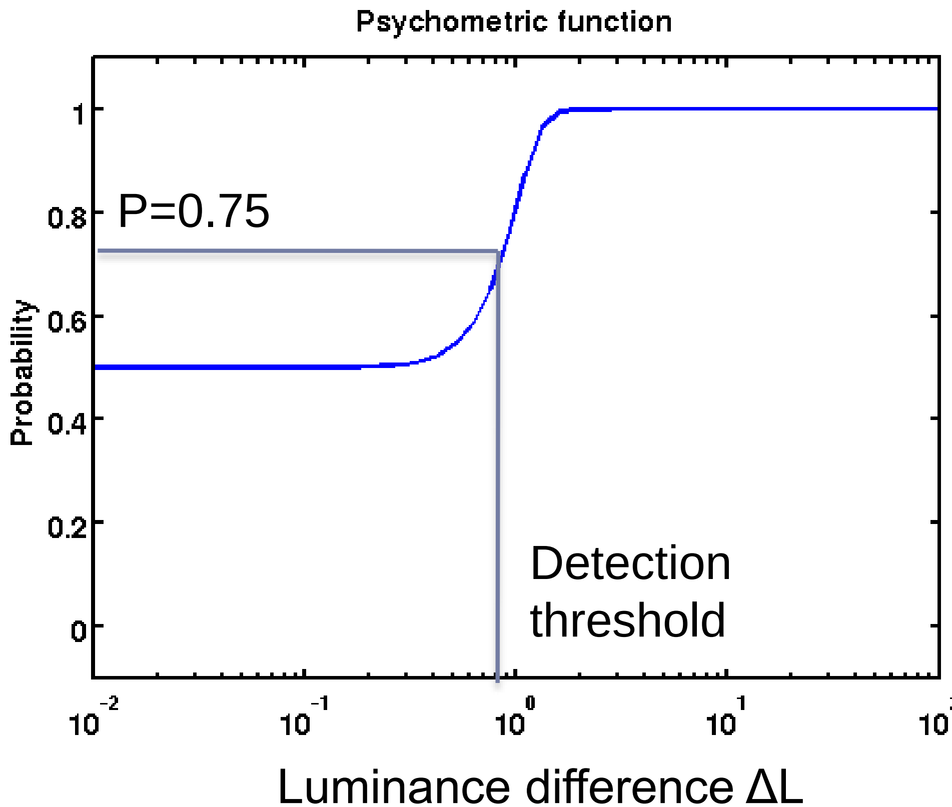


t.v.i. measurements – Blackwell 1946



Psychophysics
Threshold experiments

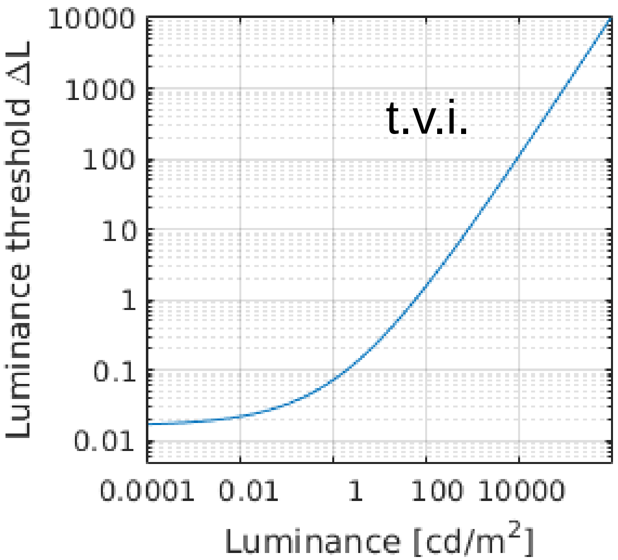




t.v.i function / c.v.i. function / Sensitivity

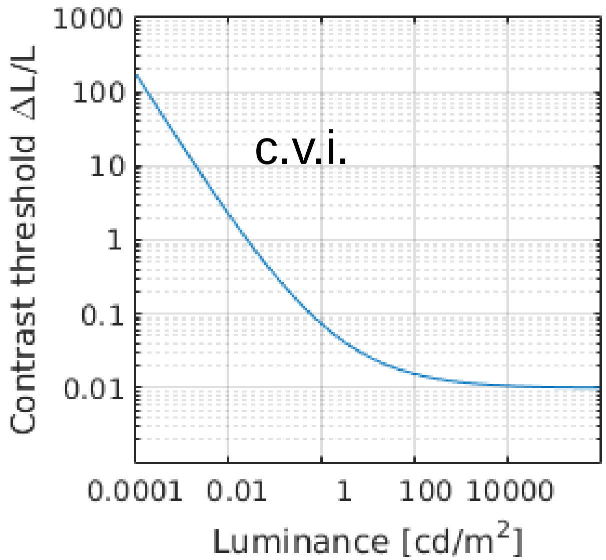
- The same data, different representation

Threshold vs. intensity



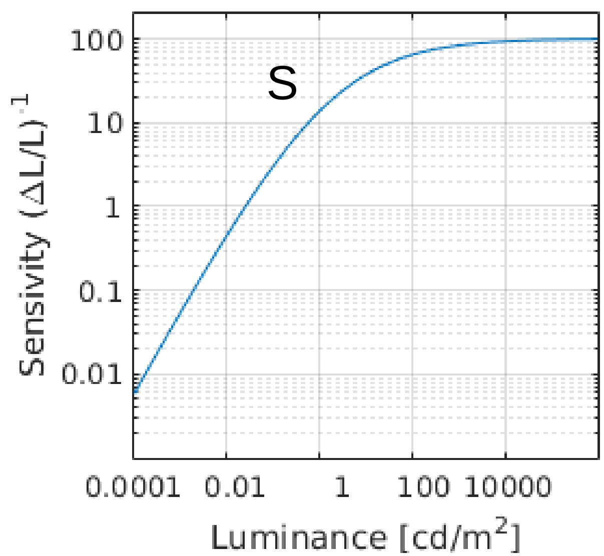
$$\Delta L = L_{disk} - L_{background}$$

Contrast vs. intensity



$$T = \frac{\Delta L}{L}$$

Sensitivity



$$S = \frac{1}{T} = \frac{L}{\Delta L}$$

Sensitivity to luminance

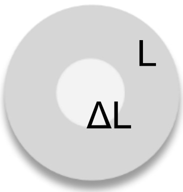
- Weber-law – the just-noticeable difference is proportional to the magnitude of a stimulus

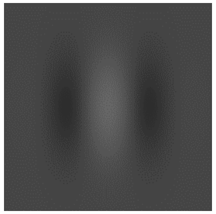
Ernst Heinrich Weber
[From wikipedia]

$$\frac{\Delta L}{L} = k$$

- ΔL: The smallest detectable luminance difference
- L: Background (adapting) luminance
- k: Constant

Typical stimuli:





Consequence of the Weber-law

- Smallest detectable difference in luminance

$$\frac{\Delta L}{L} = k$$

For k=1%

L	ΔL
100 cd/m^2	1 cd/m^2
1 cd/m^2	0.01 cd/m^2

- Adding or subtracting luminance will have different visual impact depending on the background luminance
- Unlike LDR luma values, luminance values are **not** perceptually uniform!

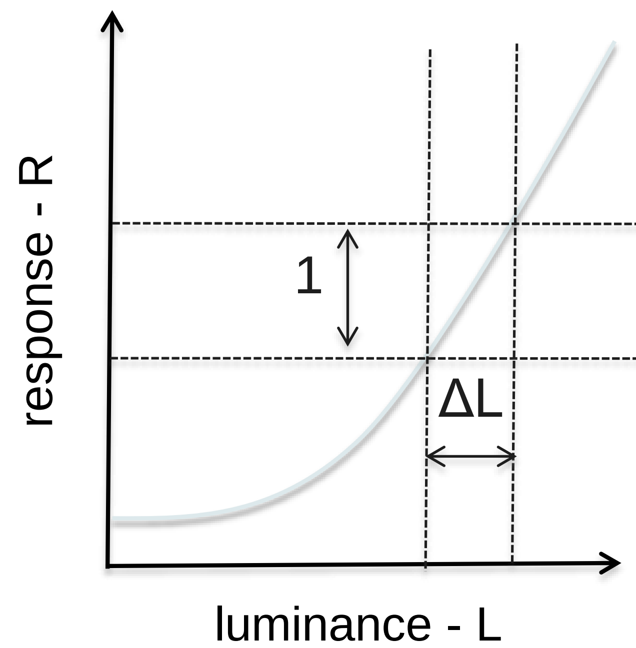
How to make luminance (more) perceptually uniform?

- Using "Fechnerian" integration

$$\frac{dR}{dl}(L) = \frac{1}{\mathrm{D}L(L)}$$

Derivative of response

Detection threshold



Luminance transducer:

$$R(L) = \int_0^L \frac{1}{\Delta L(l)} dl$$

Assuming the Weber law

$$\frac{\Delta L}{L} = k$$

- and given the luminance transducer

$$R(L) = \int_0^L \frac{1}{\Delta L(l)}\, dl$$

- the response of the visual system to light is:

$$R(L) = \int \frac{1}{kL} dL = \frac{1}{k} \ln(L) + k_1$$

Fechner law

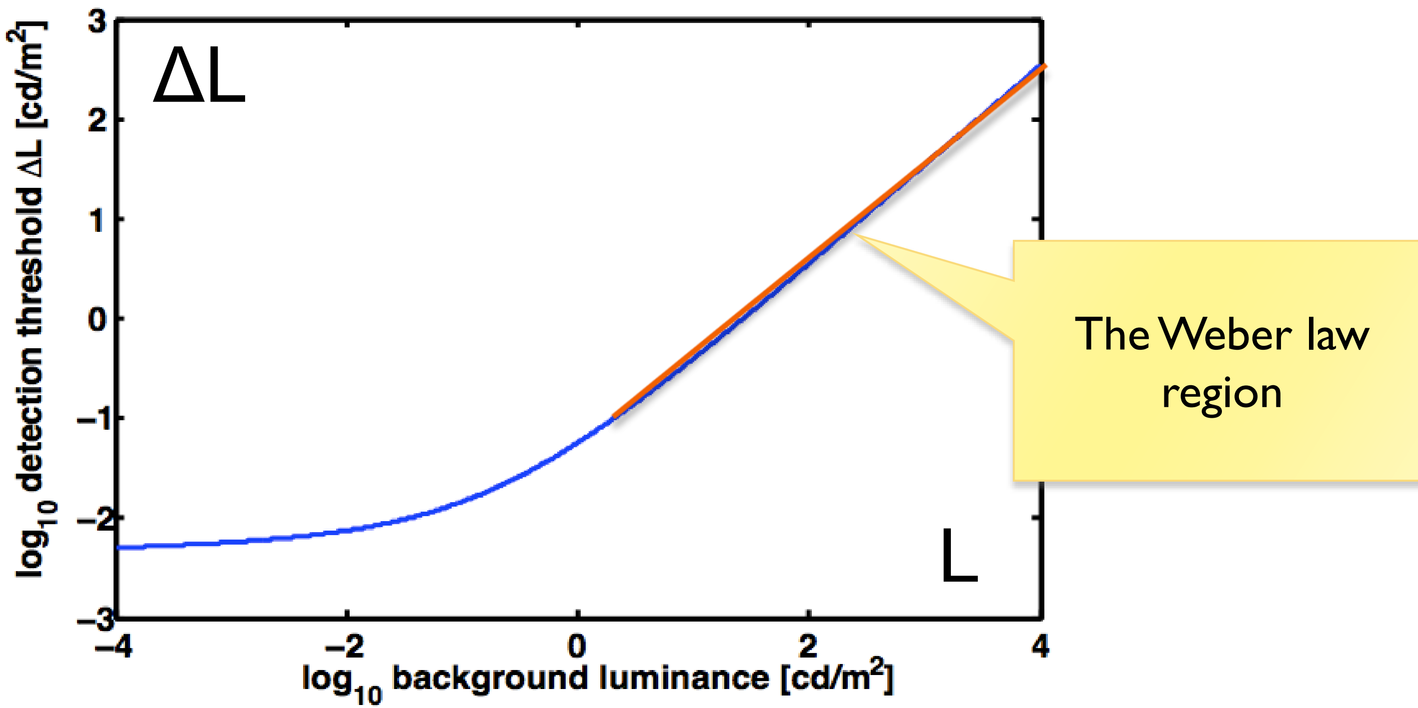
$$R(L) = a \ln(L)$$

- Response of the visual system to luminance is **approximately** logarithmic

Gustav Fechner
[From Wikipedia]

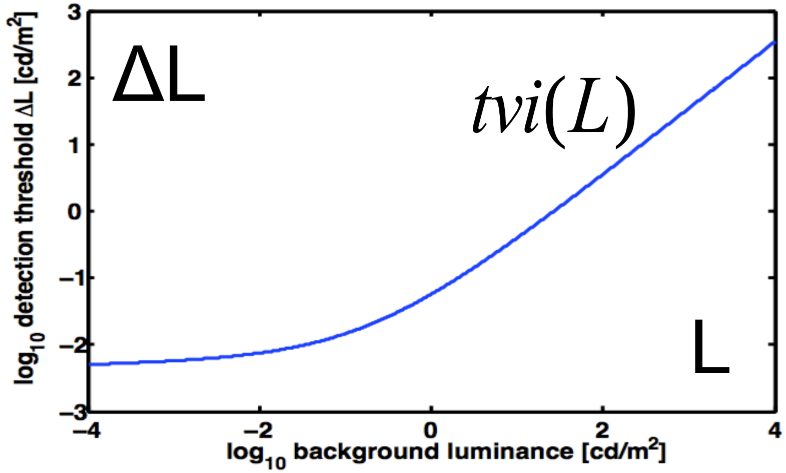
But…the Fechner law does not hold for the full luminance range

- Because the Weber law does not hold either
- Threshold vs. intensity function:



Weber-law revisited

- If we allow detection threshold to vary with luminance according to the t.v.i. function:



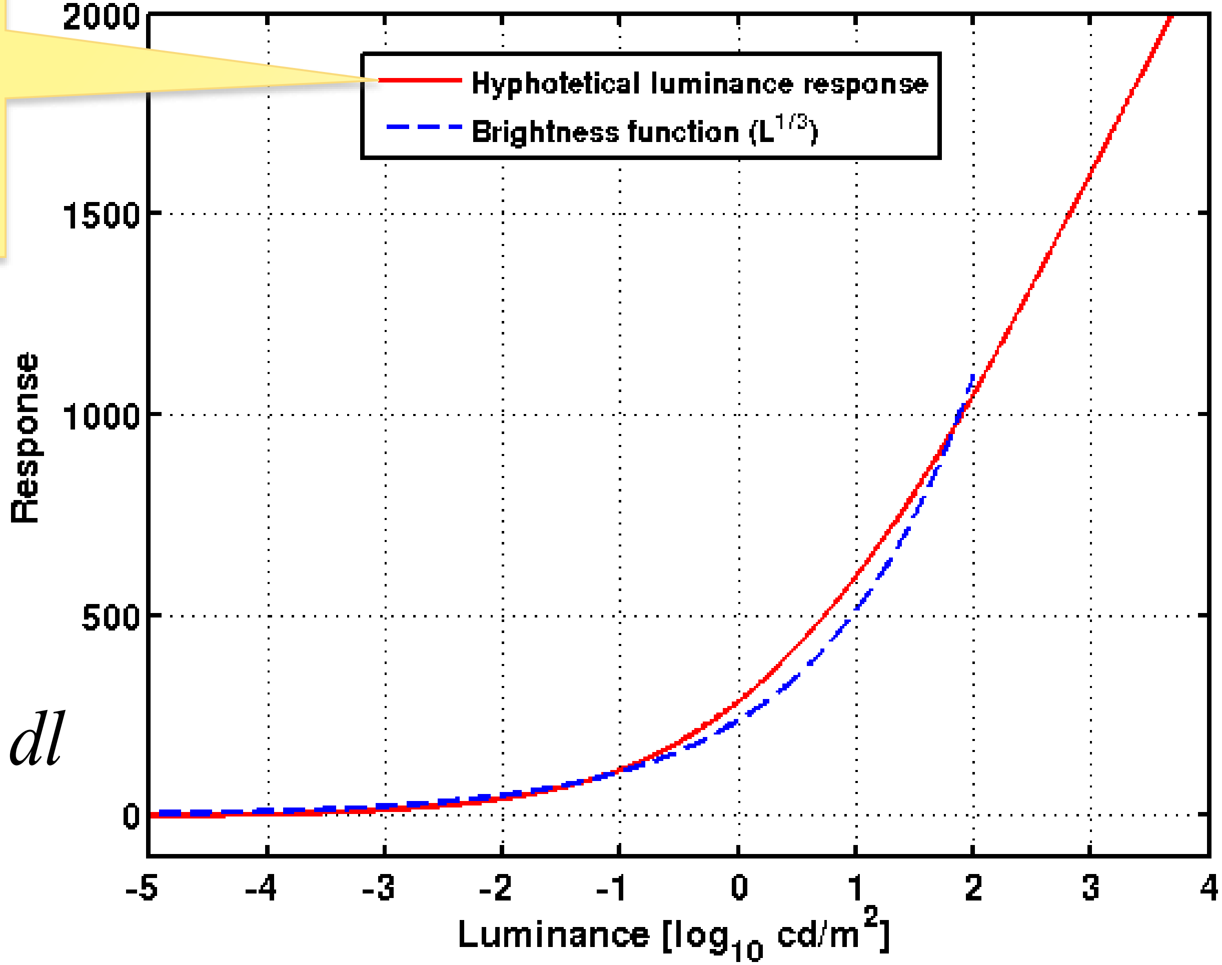
- we can get a more accurate estimate of the "response":

$$R(L) = \int_0^L \frac{1}{tvi(l)} dl$$

Fechnerian integration and Stevens' law

R(L) - function derived from the t.v.i. function

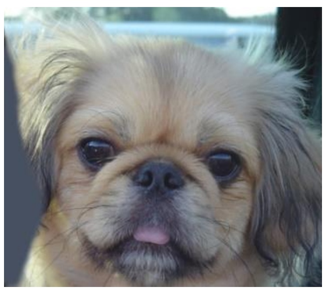
$$R(L) = \int_0^L \frac{1}{tvi(l)}\, dl$$



Applications of JND encoding – R(L)

- DICOM grayscale function
  - Function used to encode signal for medial monitors
  - 10-bit JND-scaled (just noticeable difference)
  - Equal visibility of gray levels
- HDMI 2.0a (HDR10)
  - PQ (Perceptual Quantizer) encoding
  - Dolby Vision
  - To encode pixels for high dynamic range images and video

DOLBY
VISION
The Future of Vision

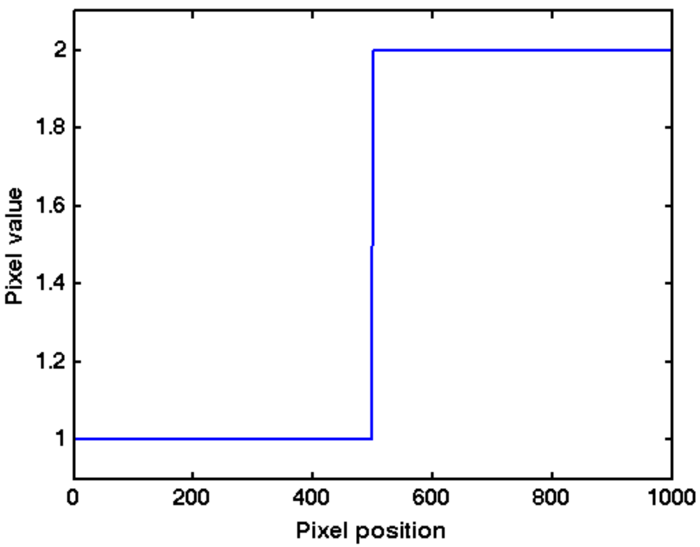



© 2006 Aude Oliva & Antonio Torralba

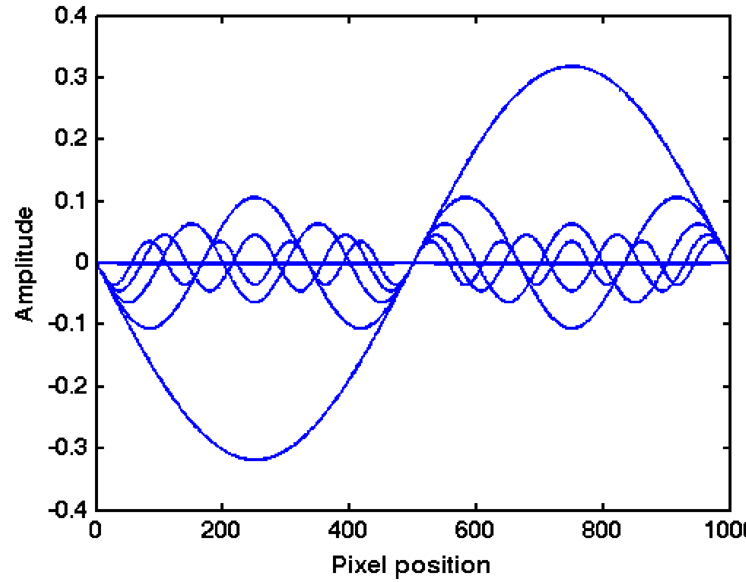
Spatial contrast sensitivity

Fourier analysis

- Every N-dimensional function (including images) can be represented as a sum of sinusoidal waves of different frequency and phase



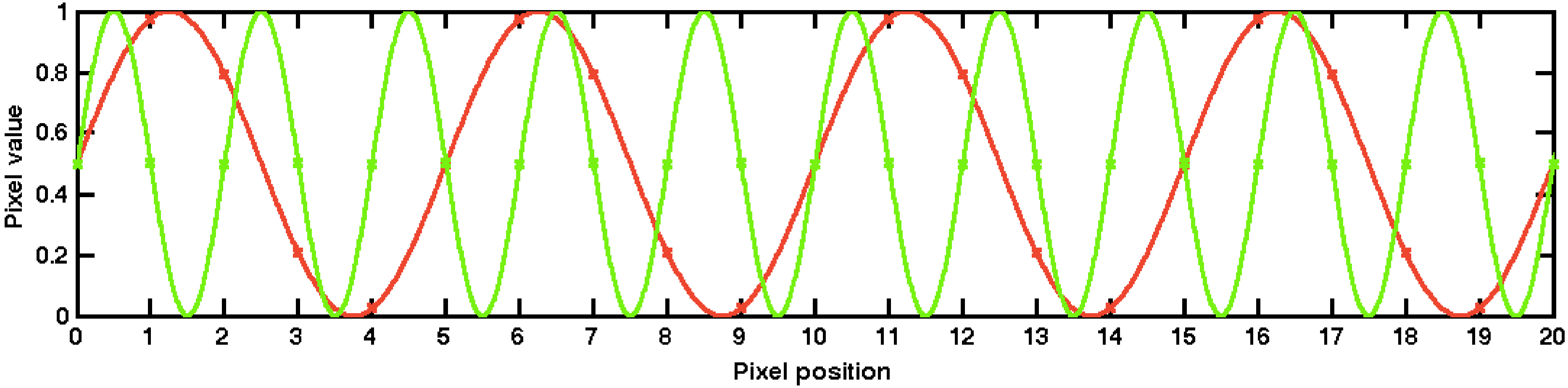
=å



- Think of "equalizer" in audio software, which manipulates each frequency

Spatial frequency in images

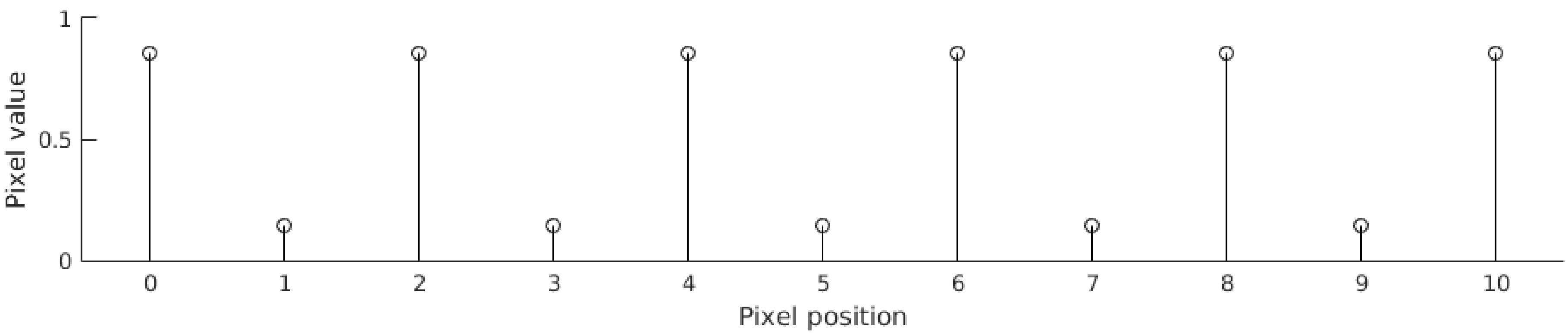
- Image space units: cycles per sample (or cycles per pixel)



  - What are the screen-space frequencies of the red and green sinusoid?
- The visual system units: cycles per degree
  - If the angular resolution of the viewed image is 55 pixels per degree, what is the frequency of the sinusoids in cycles per degree?

Nyquist frequency

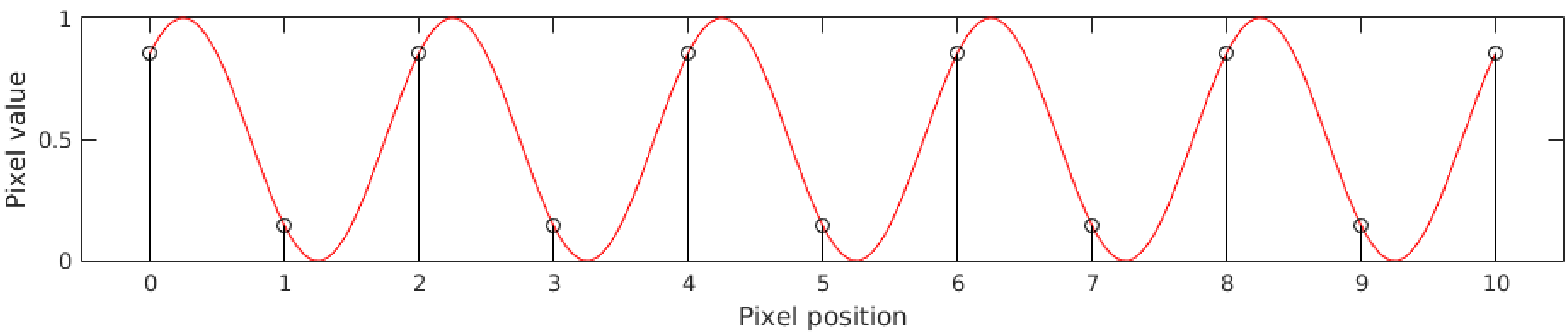
- Sampling density restricts the highest spatial frequency signal that can be (uniquely) reconstructed
  - Sampling density – how many pixels per image/visual angle/…



  - Any number of sinusoids can be fitted to this set of samples
  - It is possible to fit an infinite number of sinusoids if we allow infinitely high frequency

Nyquist frequency

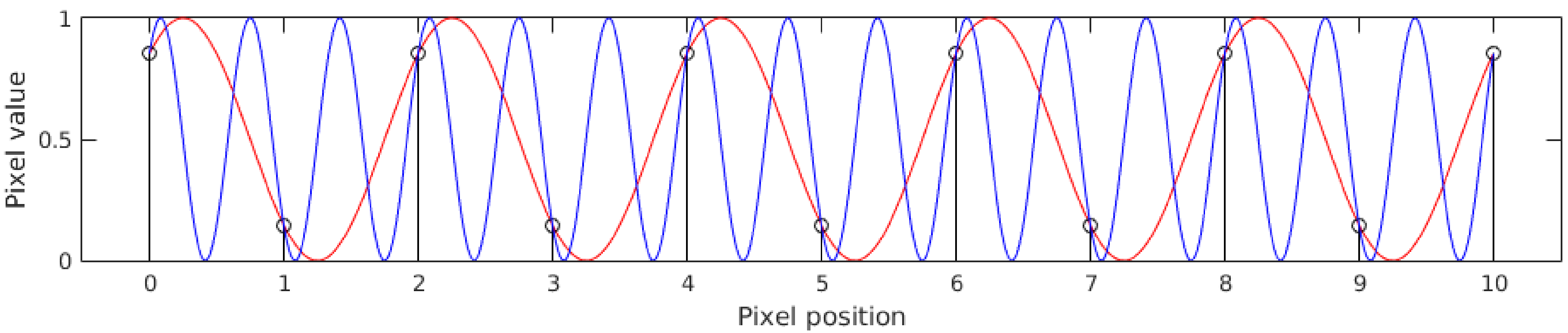
- **Sampling density restricts the highest spatial frequency signal that can be (uniquely) reconstructed**
  - Sampling density – how many pixels per image/visual angle/…



  - Any number of sinusoids can be fitted to this set of samples
  - It is possible to fit an infinite number of sinusoids if we allow infinitely high frequency

Nyquist frequency

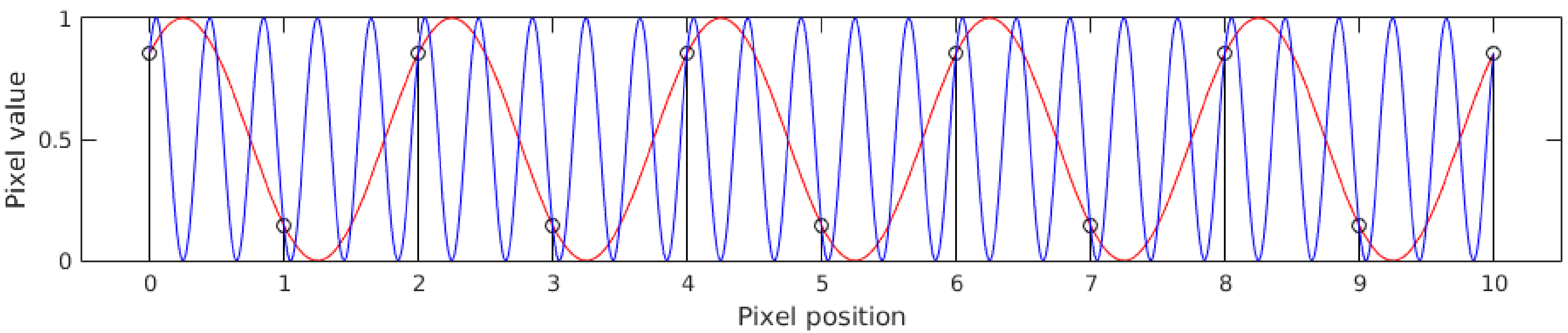
- Sampling density restricts the highest spatial frequency signal that can be (uniquely) reconstructed
  - Sampling density – how many pixels per image/visual angle/…



  - Any number of sinusoids can be fitted to this set of samples
  - It is possible to fit an infinite number of sinusoids if we allow infinitely high frequency

Nyquist frequency

- Sampling density restricts the highest spatial frequency signal that can be (uniquely) reconstructed
  - Sampling density – how many pixels per image/visual angle/…



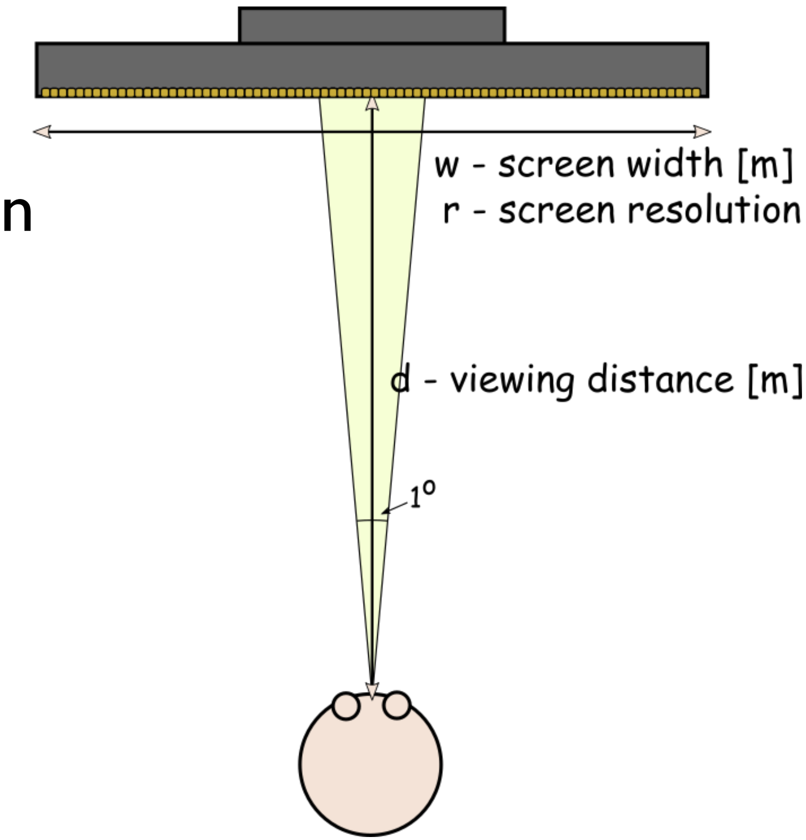
  - Any number of sinusoids can be fitted to this set of samples
  - It is possible to fit an infinite number of sinusoids if we allow infinitely high frequency

Nyquist frequency / aliasing

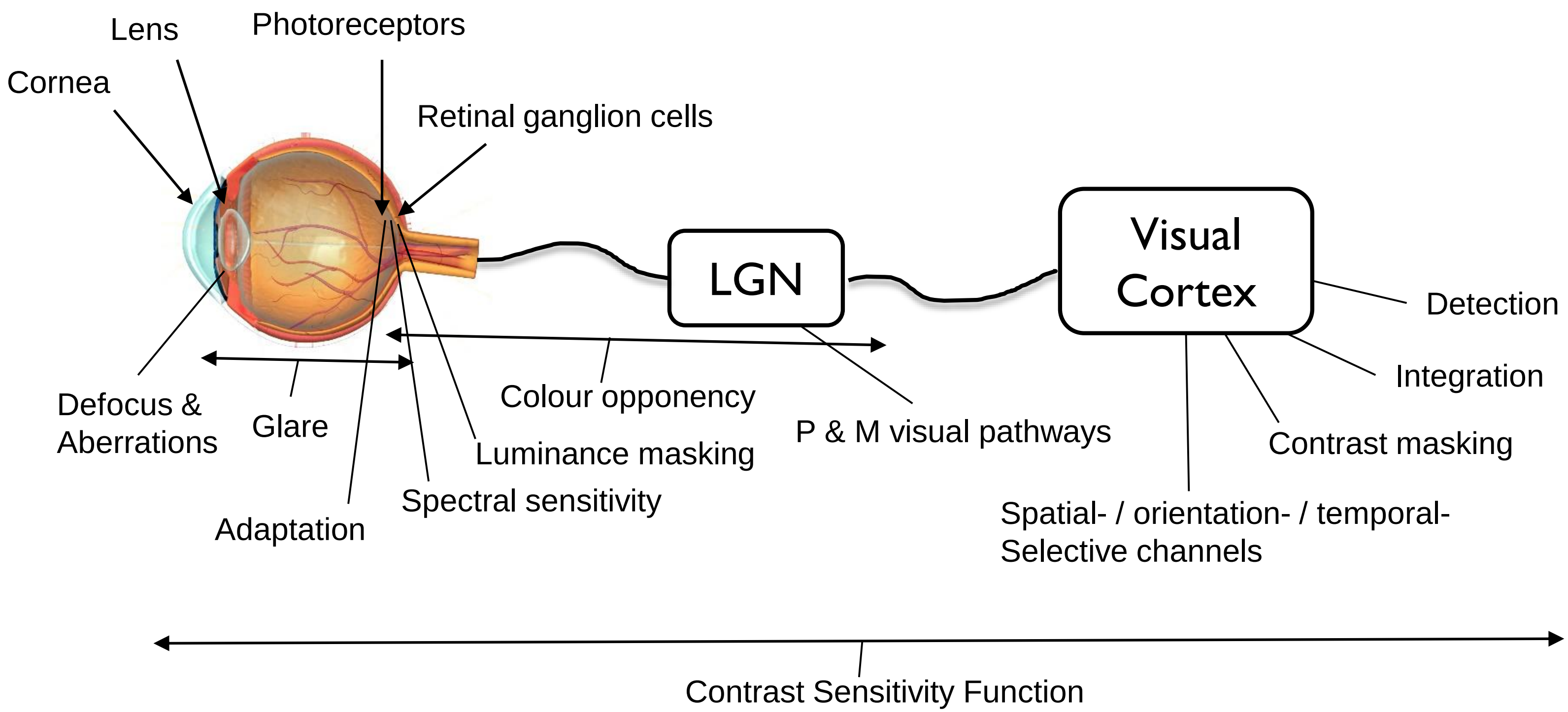
- Nuquist frequency is the highest frequency that can be represented by a discrete set of uniform samples (pixels)
- Nuquist frequency = 0.5 sampling rate
  - For audio
    - If the sampling rate is 44100 samples per second (audio CD), then the Nyquist frequency is 22050 Hz
  - For images (visual degrees)
    - If the sampling rate is 60 pixels per degree, then the Nyquist frequency is 30 cycles per degree
- When resampling an image to lower resolution, the frequency content above the Nyquist frequency needs to be removed (reduced in practice)
  - Otherwise **aliasing** is visible

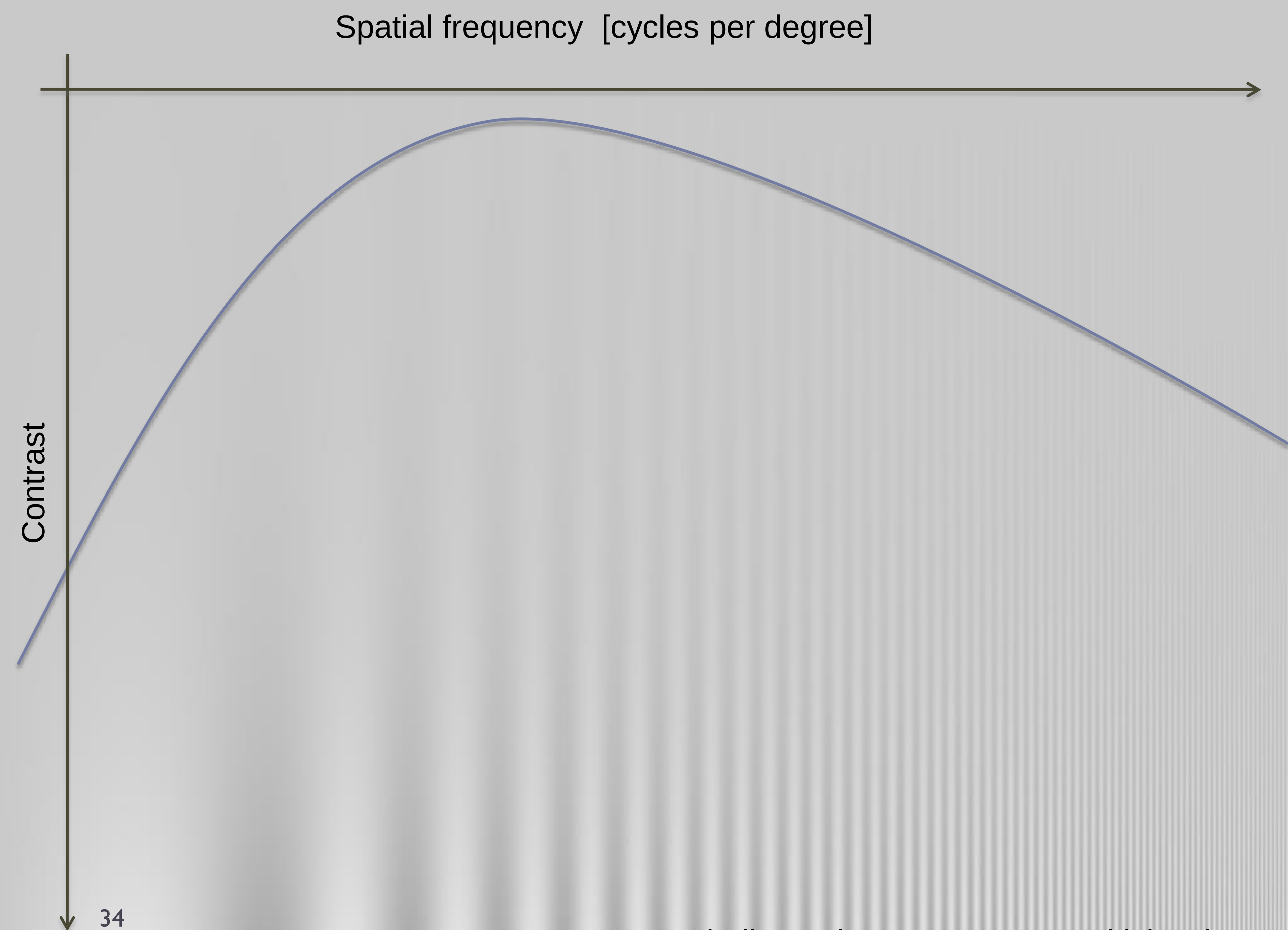
Resolution and sampling rate

- Pixels per inch [ppi]
  - Does not account for vision
- The visual resolution depends on
  - screen size
  - screen resolution
  - viewing distance
- The right measure
  - Pixels per visual degree [ppd]
  - In frequency space
    - Cycles per visual degree [cpd]
    - Nyquist frequency [cpd] = ½ of [ppd]

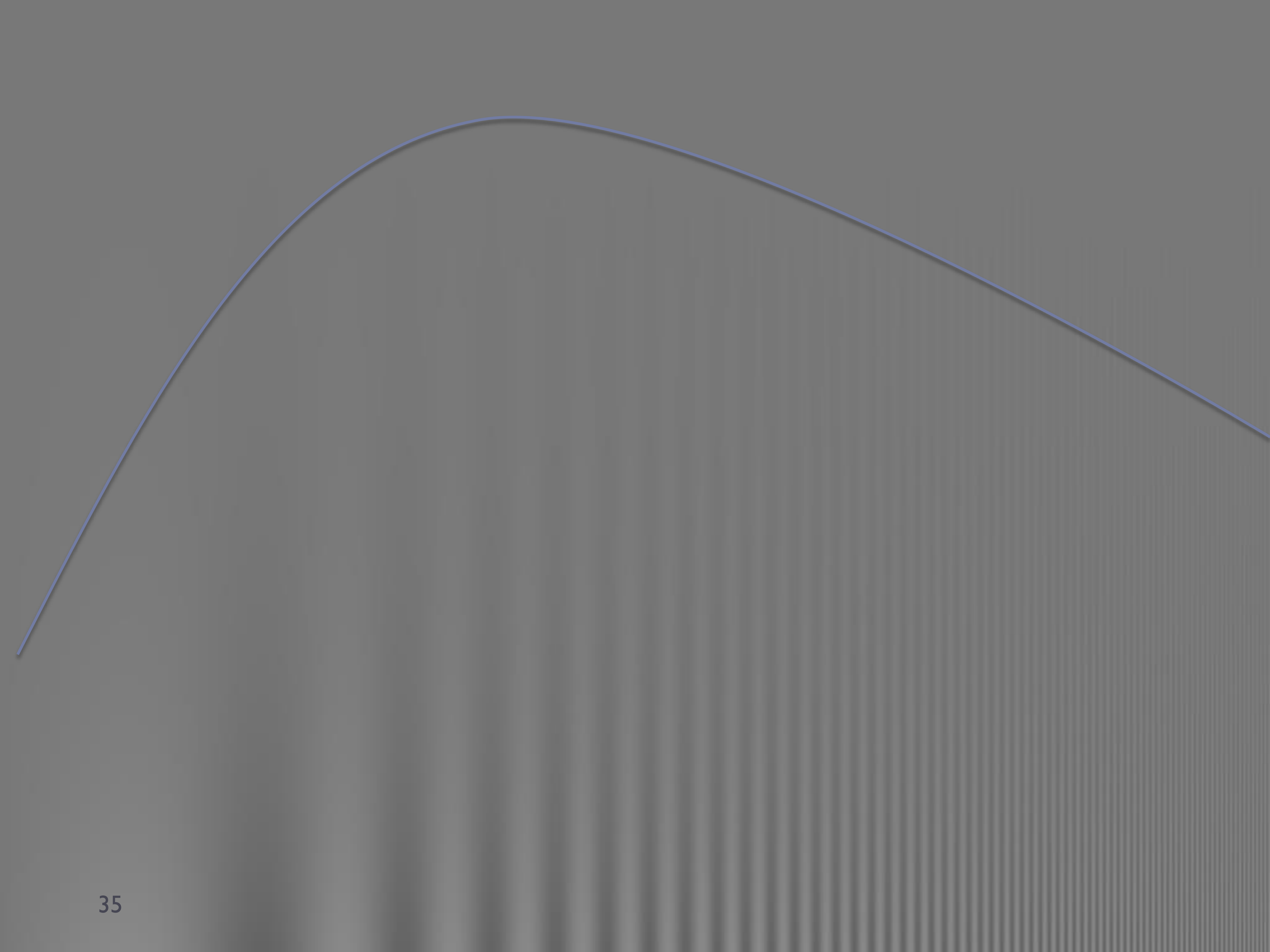


Modeling visual system

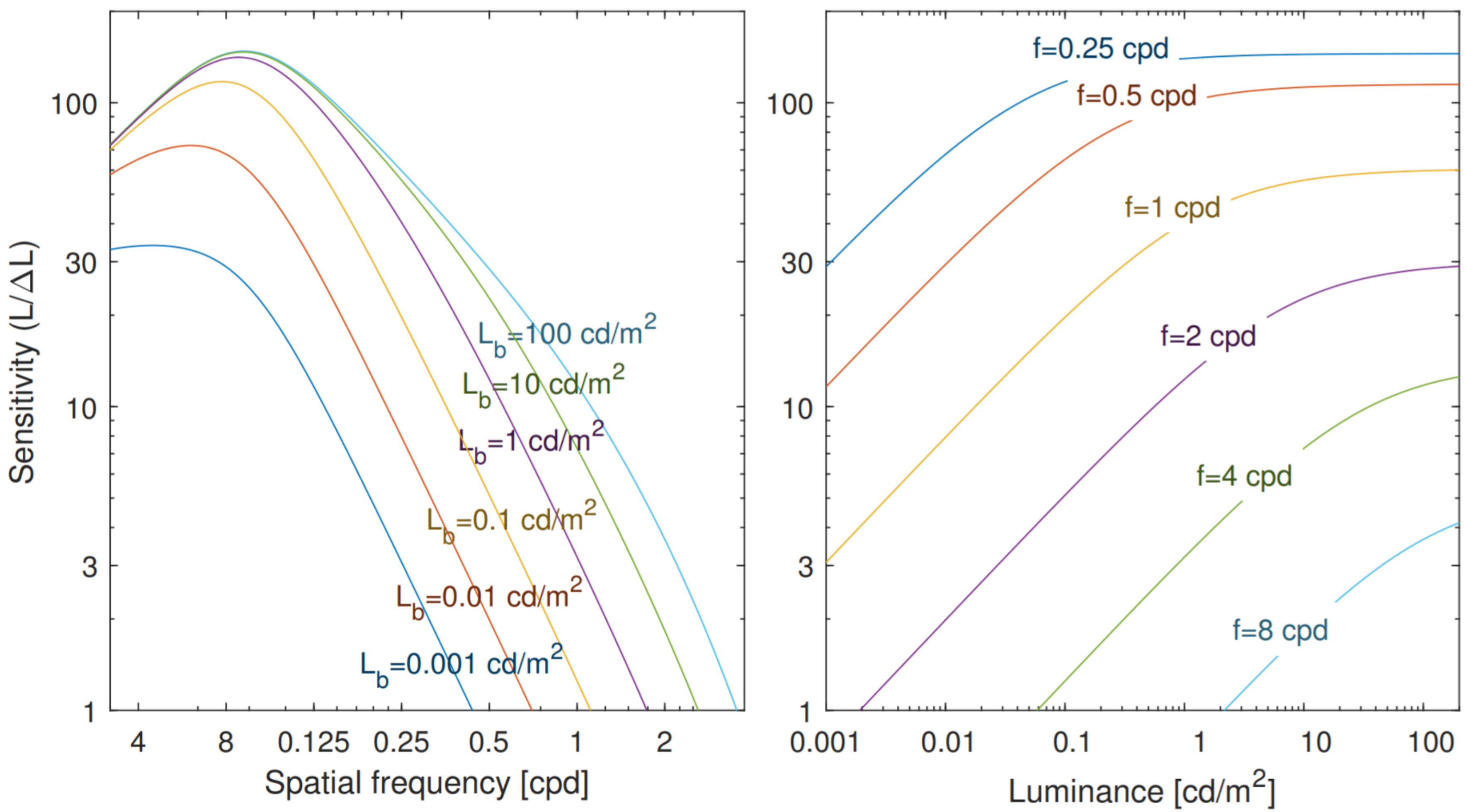




Campbell & Robson contrast sensitivity chart



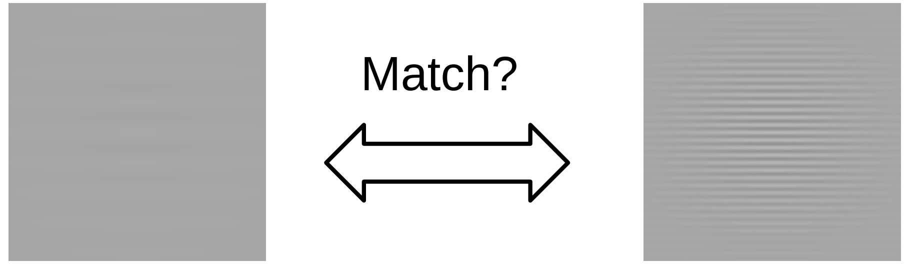
CSF as a function of spatial frequency and background luminance

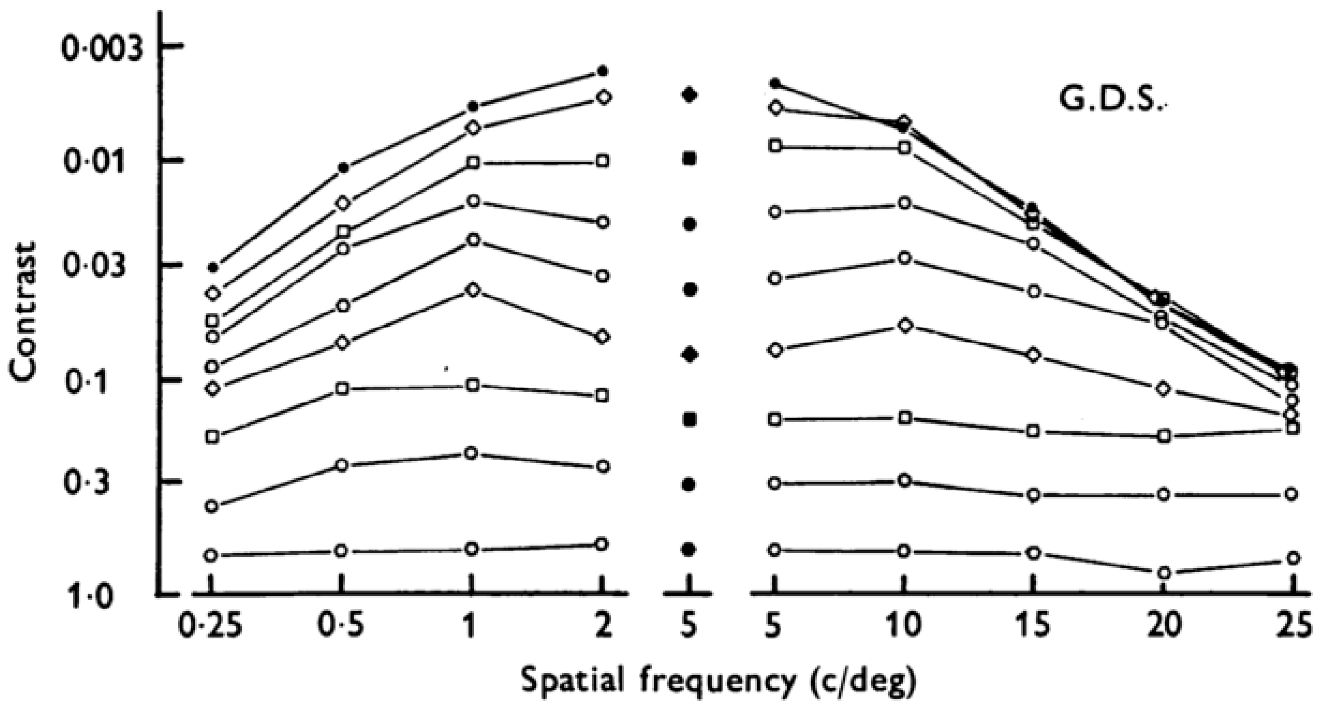


Contrast constancy

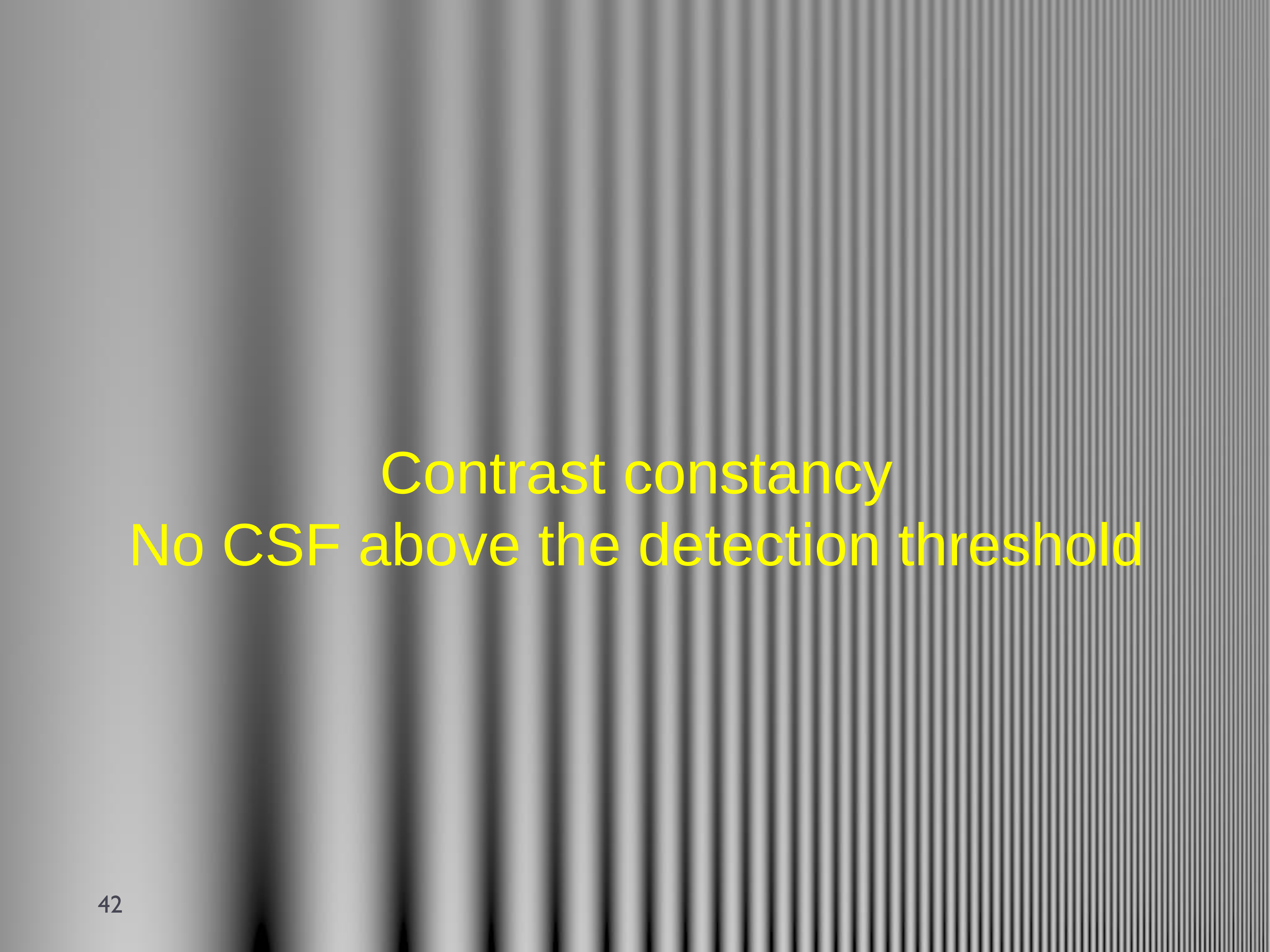
Contrast constancy

Experiment: Adjust the amplitude of one sinusoidal grating until it matches the perceived magnitude of another sinusoidal grating.



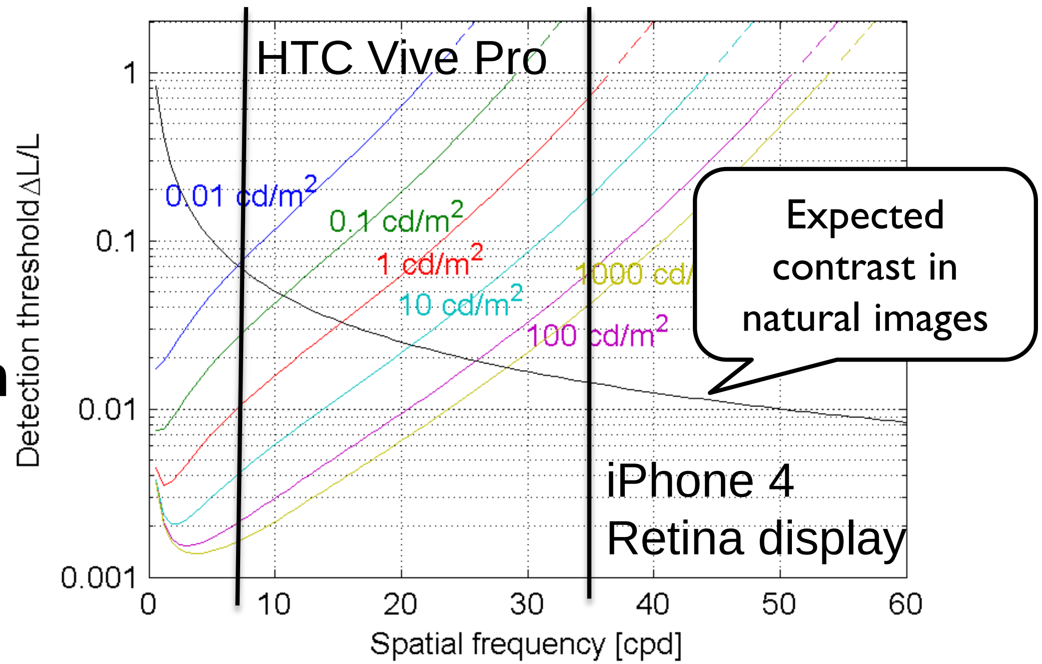


From: Georgeson and Sullivan. 1975. J. Phsysio.


Contrast constancy
No CSF above the detection threshold

CSF and the resolution

- CSF plotted as the detection contrast
$$\frac{\Delta L}{L_b} = S^{-1}$$
- The contrast below each line is invisible
- Maximum perceivable resolution depends on luminance

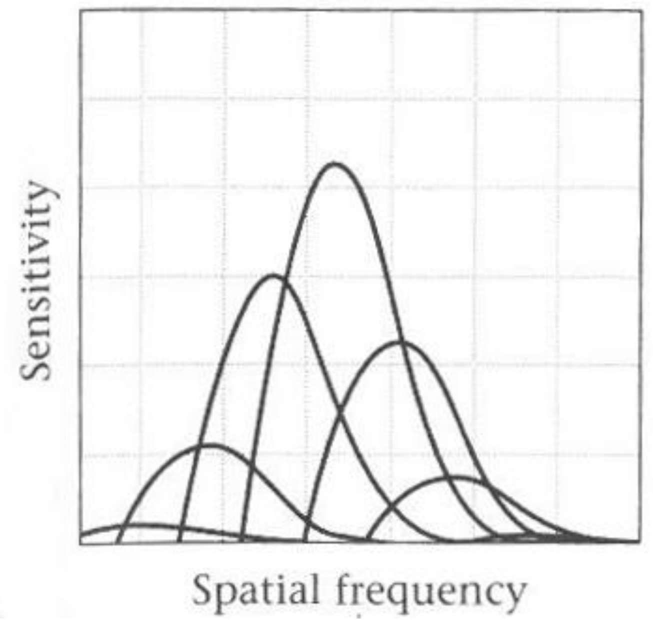


CSF models:
Barten, P. G. J. (2004). https://doi.org/10.1117/12.537476

Multi-resolution models

Spatial-frequency selective channels

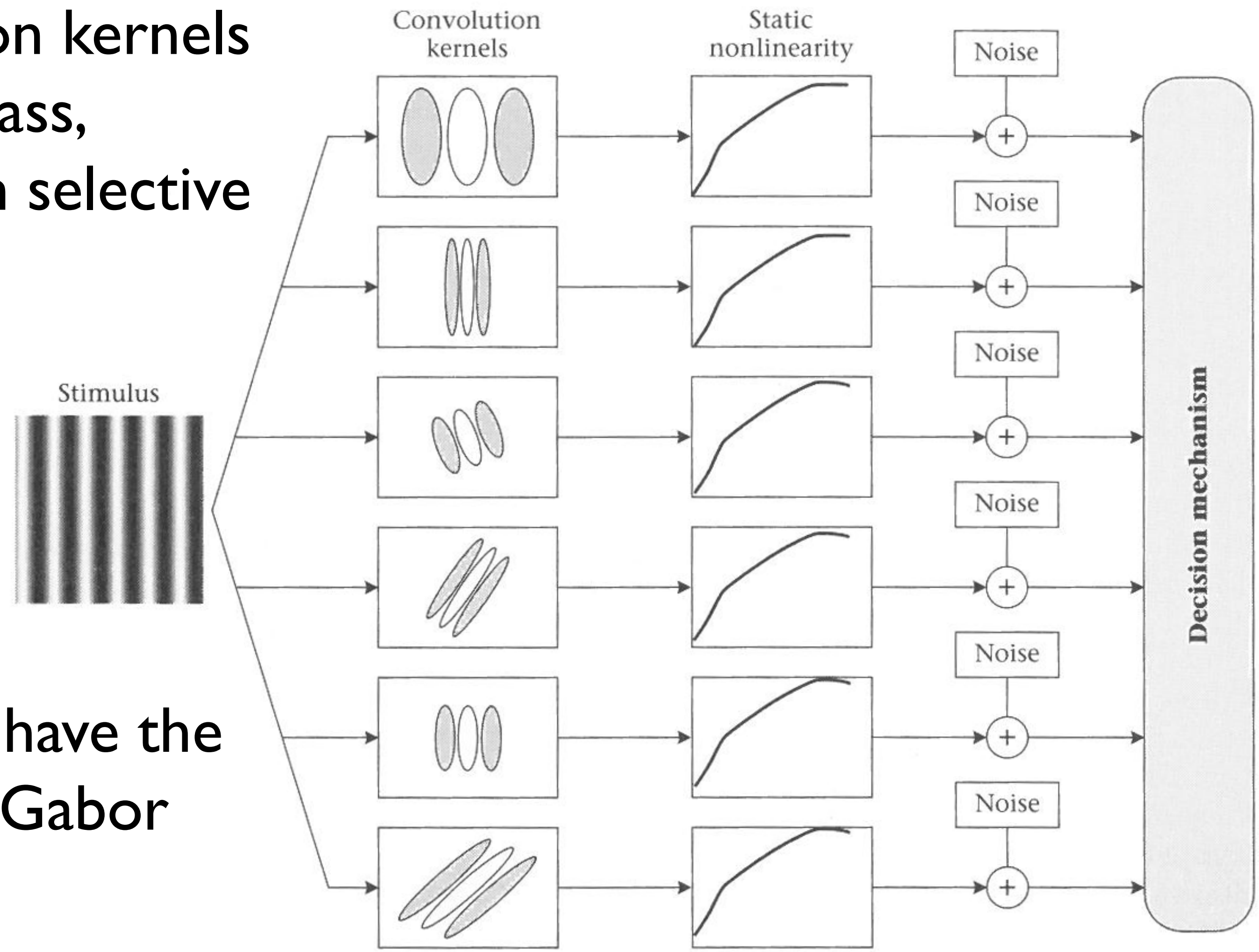
- The visual information is decomposed in the visual cortex into multiple channels
  - The channels are selective to spatial frequency, temporal frequency and orientation
  - Each channel is affected by different „noise” level
  - The CSF is the net result of information being passed in noise-affected visual channels



From: Wandell, 1995

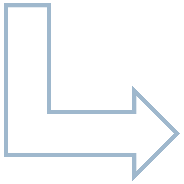
Multi-resolution visual model

- Convolution kernels are band-pass, orientation selective filters
- The filters have the shape of a Gabor function

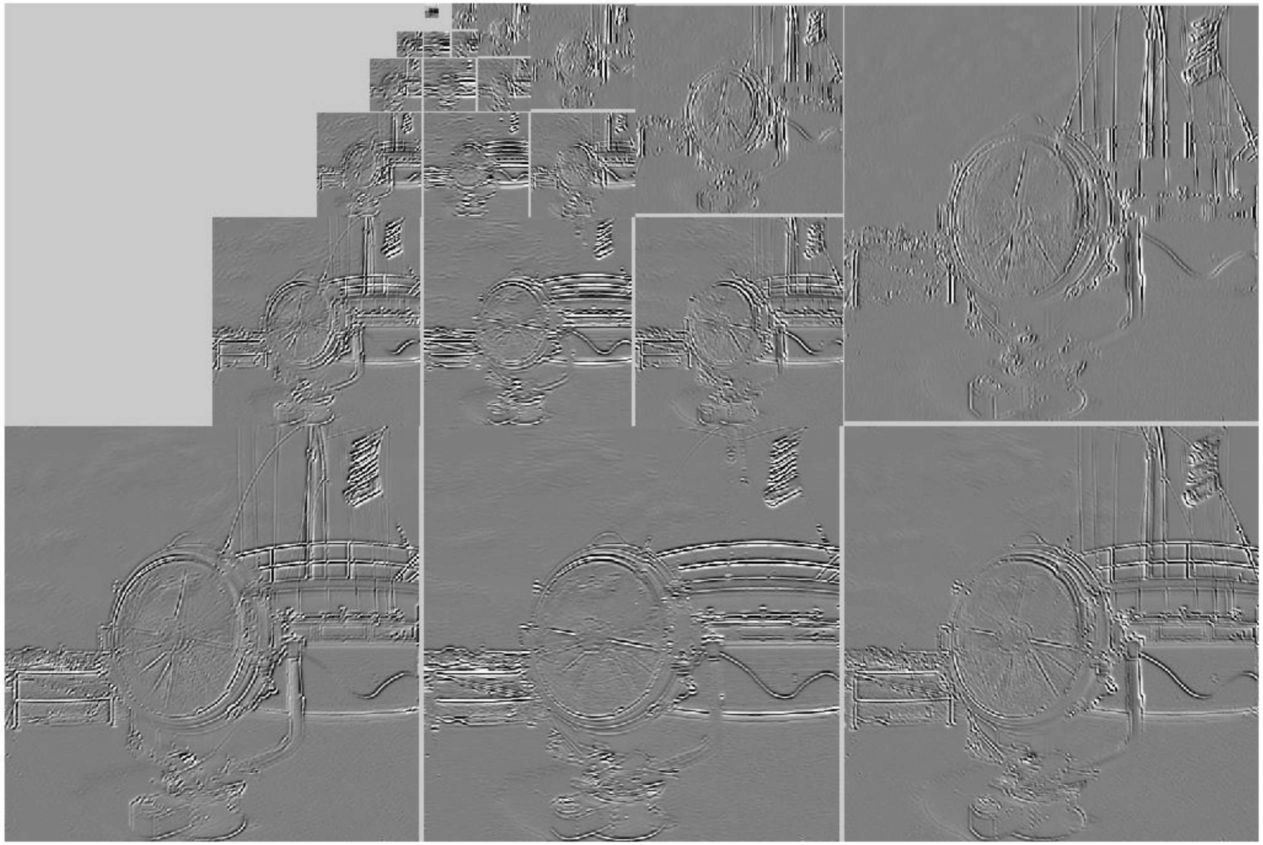


From: Wandell, 1995

Multi-scale decomposition

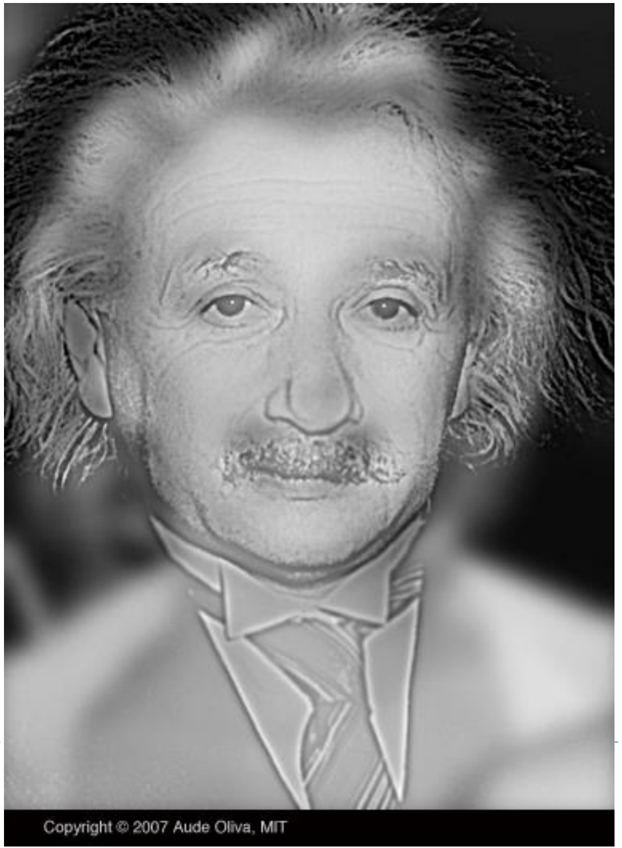


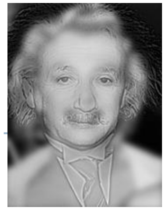
Steerable pyramid decomposition



Applications of multi-scale models

- JPEG2000
  - Wavelet decomposition
- JPEG / MPEG
  - Frequency transforms
- Image pyramids
  - Blending & stitching
  - Hybrid images



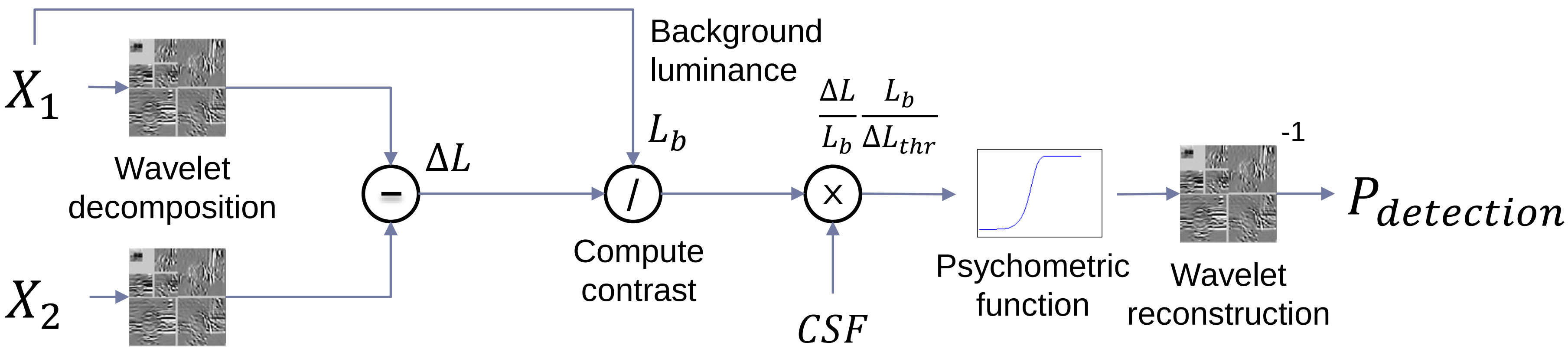


Hybrid Images by Aude Oliva
http://cvcl.mit.edu/hybrid_gallery

Predicting visible differences with CSF

- But we can use CSF to find the probability of spotting a difference beween a pair of images X_1 and X_2:

$$p(f[X_1] = f[X_2]\,|X_1, X_2, CSF)$$

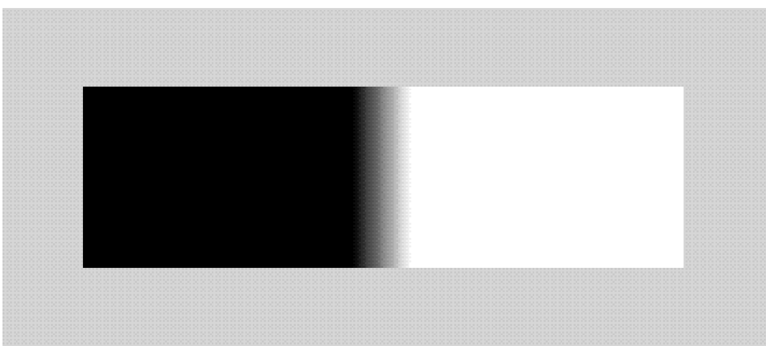


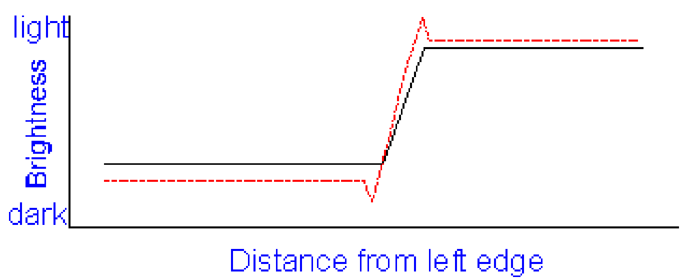
(simplified) Visual Difference Predictor

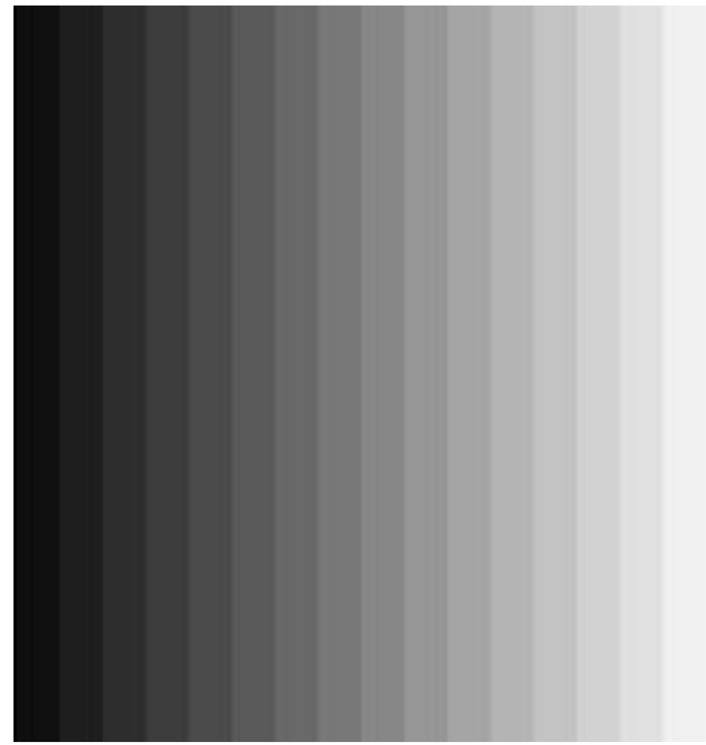
Daly, S. (1993).

Mach Bands – evidence for band-pass visual processing

- “Overshooting“ along edges
  - Extra-bright rims on bright sides
  - Extra-dark rims on dark sides
- Due to “Lateral Inhibition“



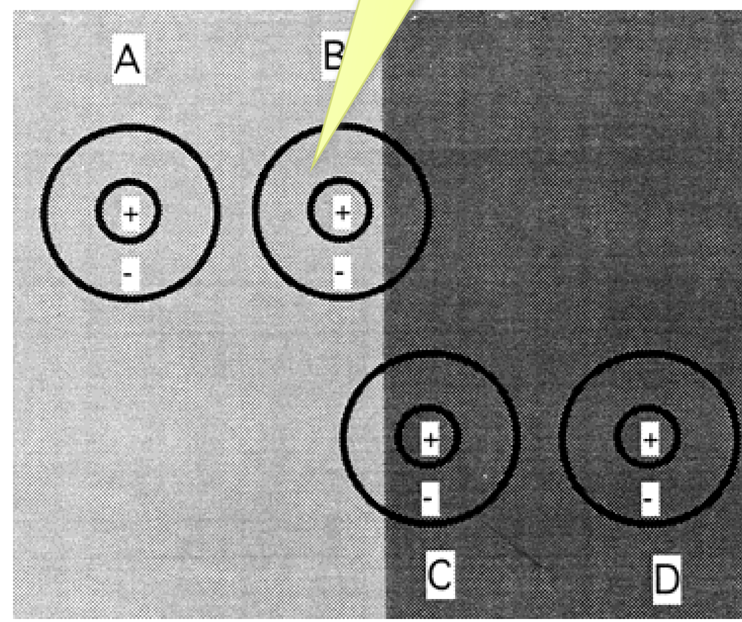


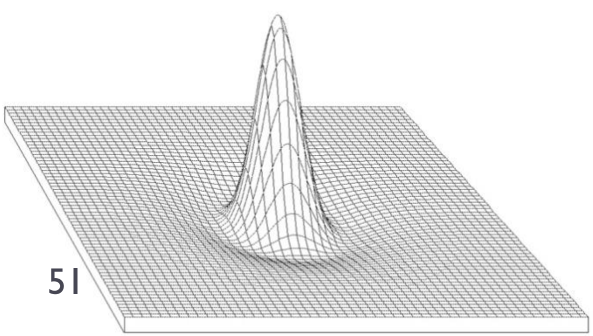


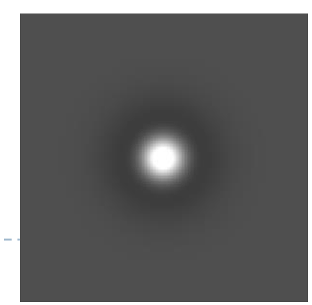
Centre-surround (Lateral Inhibition)

- "Pre-processing" step within the retina
  - Surrounding brightness level weighted negatively
    - A: high stimulus, maximal bright inhibition
    - B: high stimulus, reduced inhibition & stronger response
    - D: low stimulus, maximal inhibition
    - C: low stimulus, increased inhibition & weaker response

Center-surround receptive fields (groups of photoreceptors)

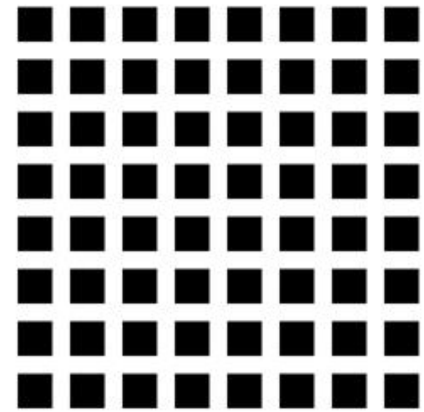


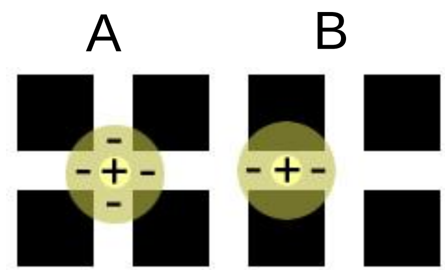


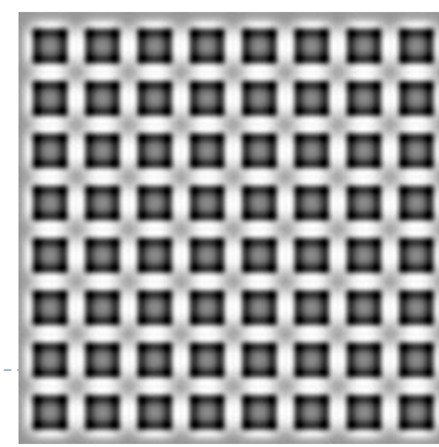


Centre-surround: Hermann Grid

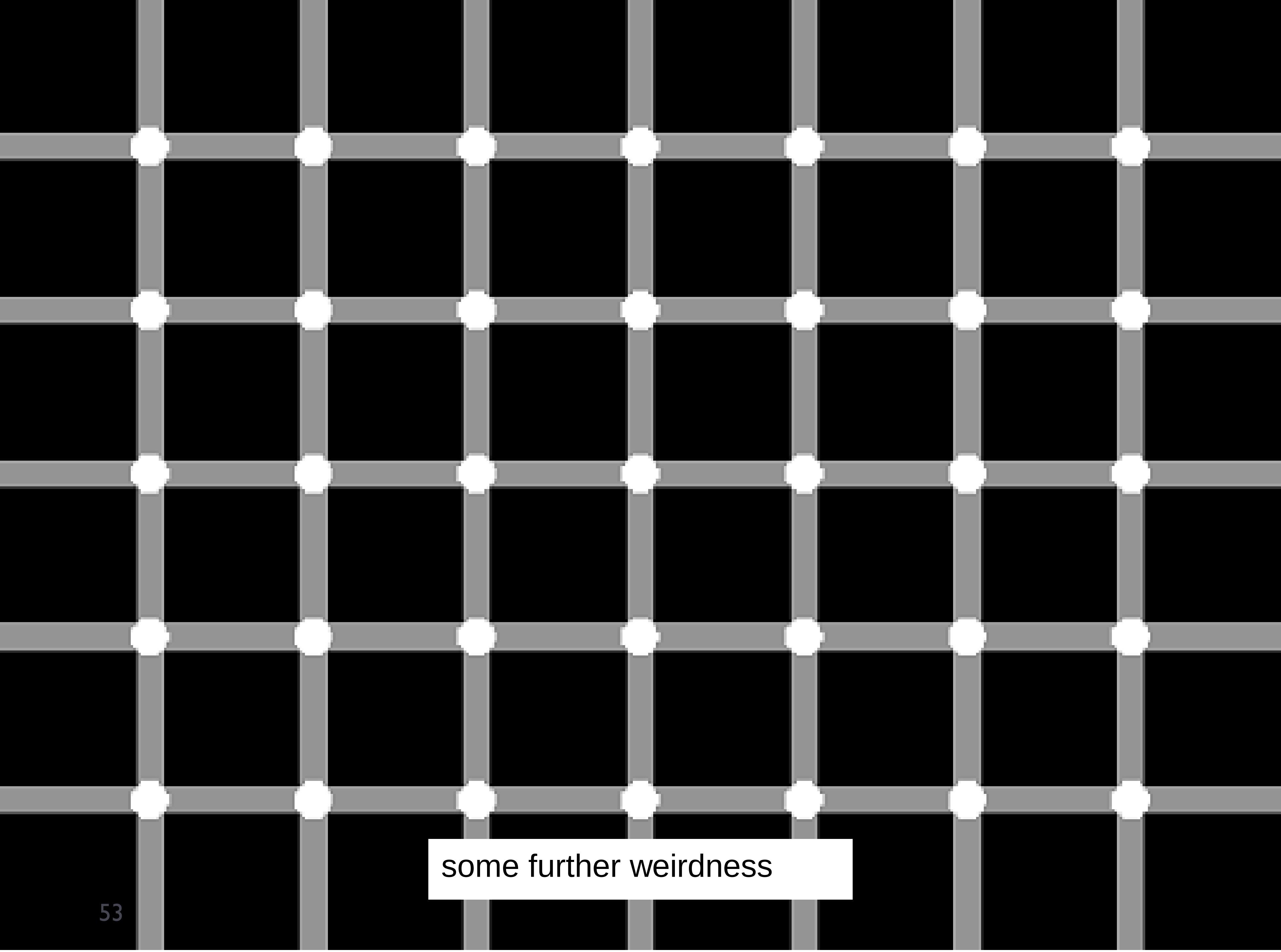
- Dark dots at crossings
- Explanation
  - Crossings (A)
    - More surround stimulation (more bright area)
    - ⇒ Less inhibition
    - ⇒ Weaker response
  - Streets (B)
    - Less surround stimulation
    - ⇒ More inhibition
    - ⇒ Greater response
- Simulation
  - Darker at crossings, brighter in streets
  - Appears more steady
  - What if reversed ?





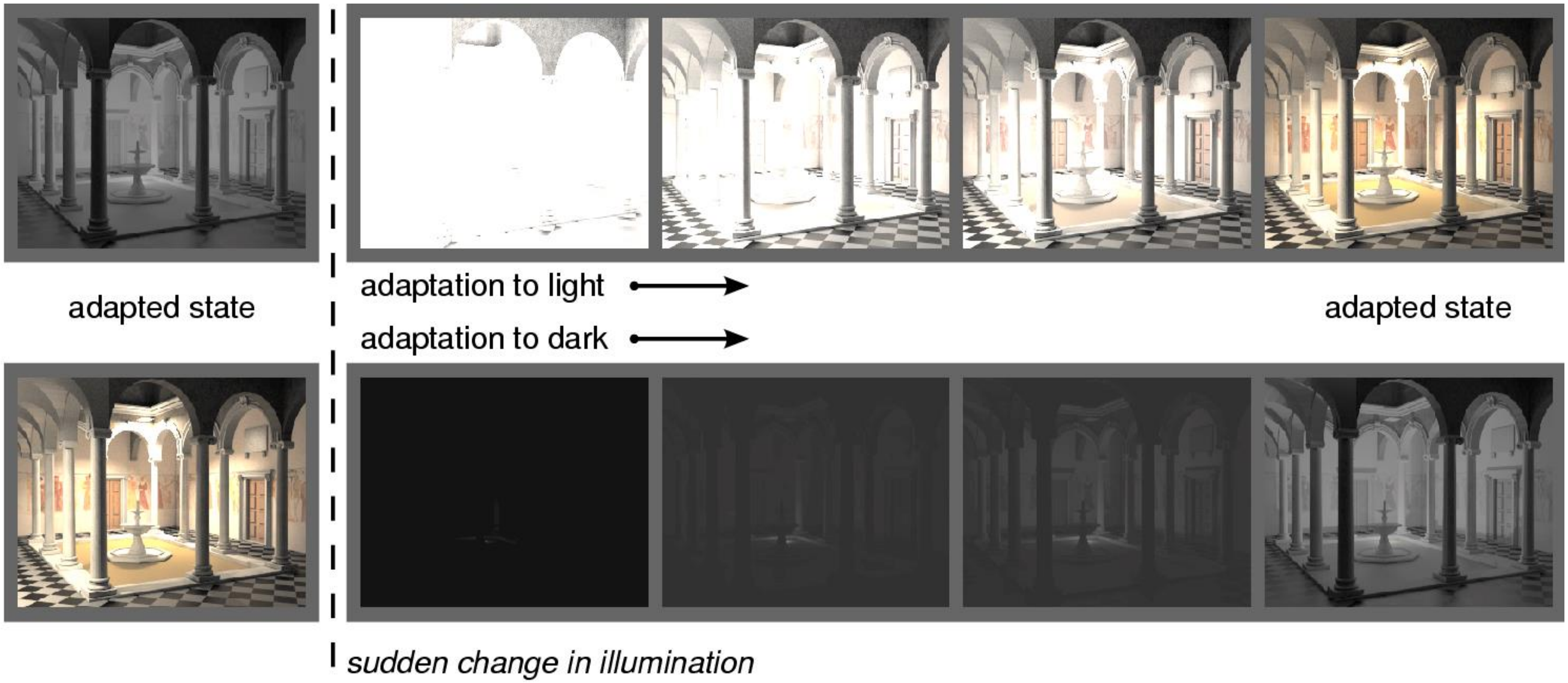


Simulation


some further weirdness

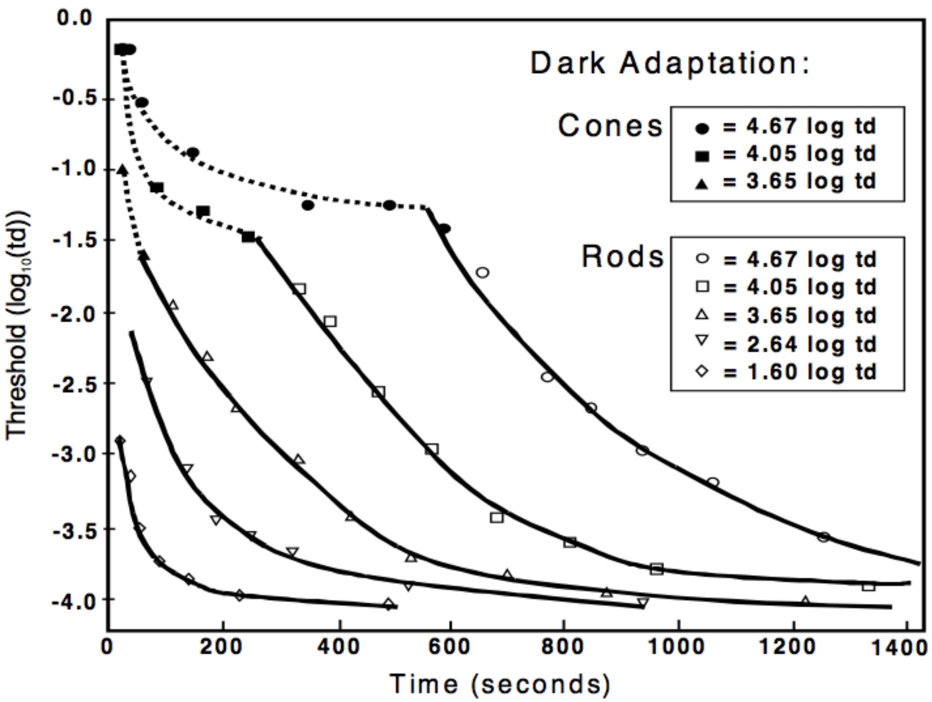
Light and dark adaptation

Light and dark adaptation

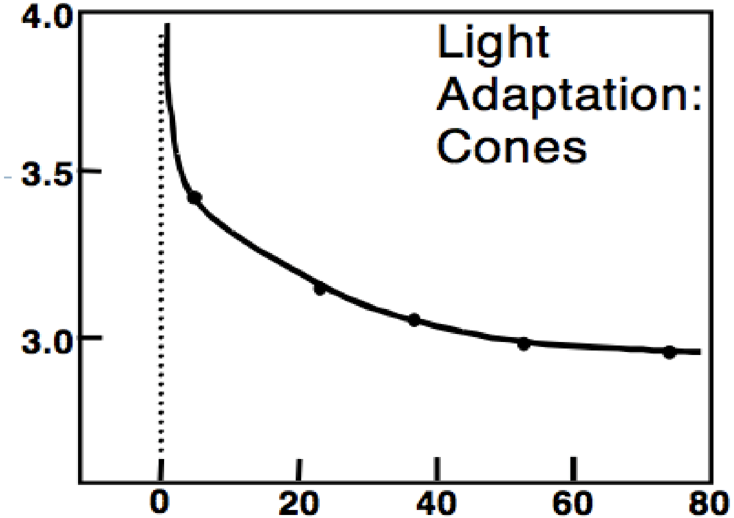


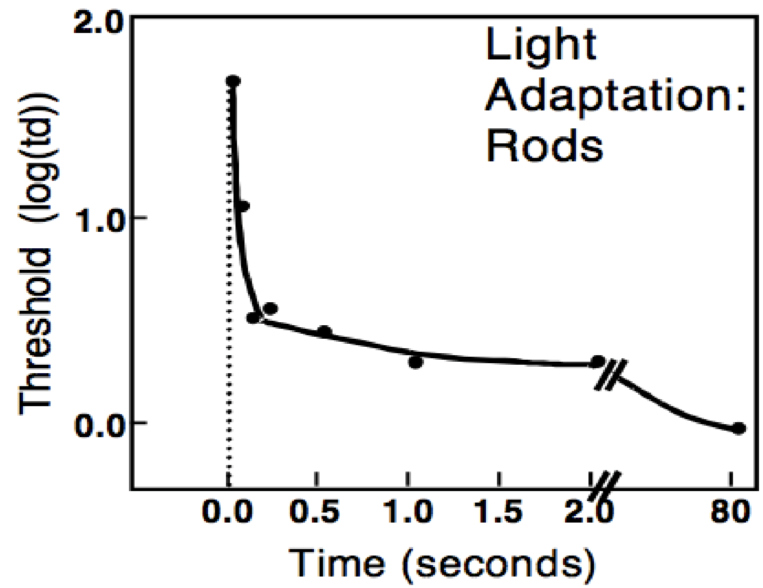
- Light adaptation: from dark to bright
- Dark adaptation: from bright to dark (much slower)

Time-course of adaptation



Bright -> Dark





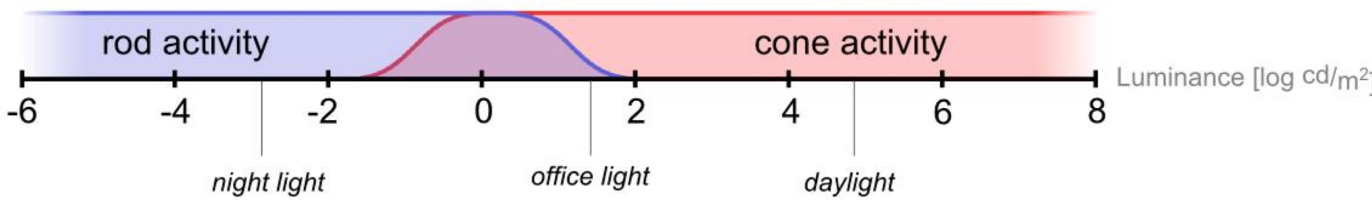
Dark -> Bright

Temporal adaptation mechanisms

- Bleaching & recovery of photopigment
  - Slow assymetric (light -> dark, dark -> light)
  - Reaction times (1-1000 sec)
  - Separate time-course for rods and cones
- Neural adaptation
  - Fast
  - Approx. symmetric reaction times (10-3000 ms)
- Pupil
  - Diameter varies between 3 and 8 mm
  - About 1:7 variation in retinal illumunation

Night and daylight vision

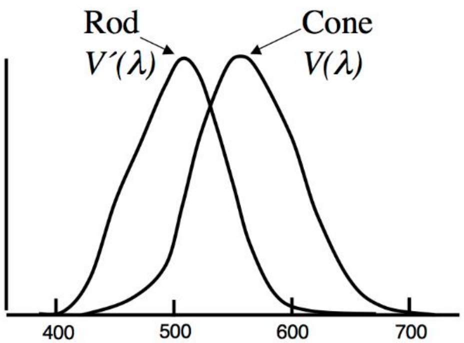
Vision mode: SCOTOPIC MESOPIC PHOTOPIC



Mode properties: monochromatic vision
limited visual acuity

good color perception
good visual acuity

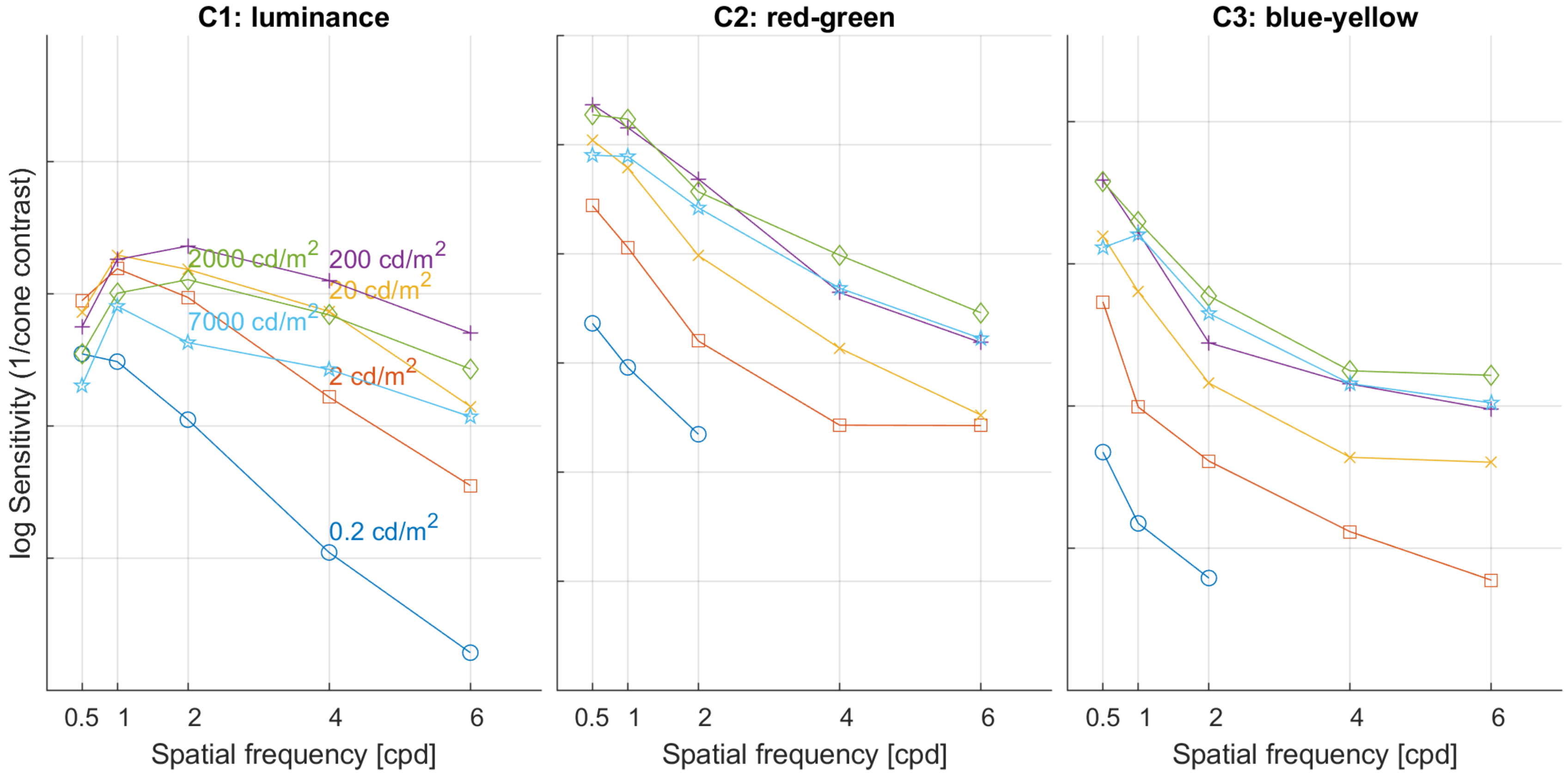
Luminous efficiency



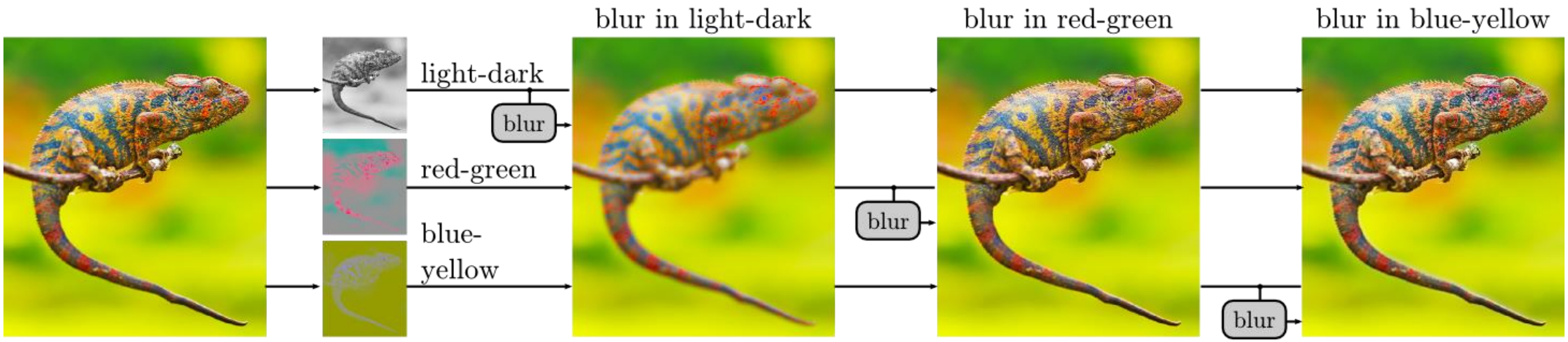
Spatial colour vision

Spatio-chromatic CSF

Color CSF across the luminance range



Visibility of blur

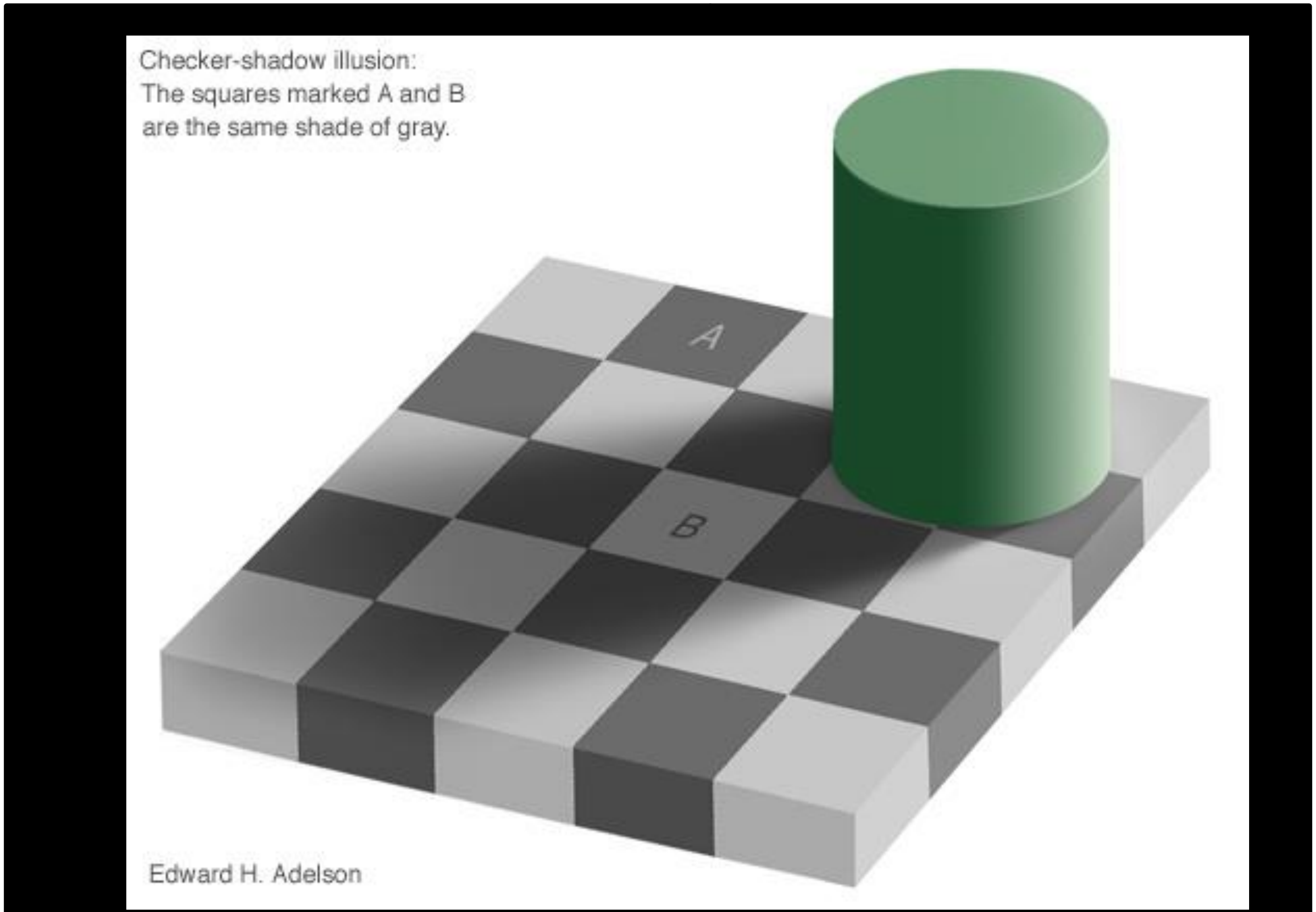


- The same amount of blur was introduced into light-dark, red-green and blue-yellow colour opponent channels
- The blur is only visible in light-dark channel
- This property is used in image and video compression
  - Sub-sampling of colour channels (4:2:1)

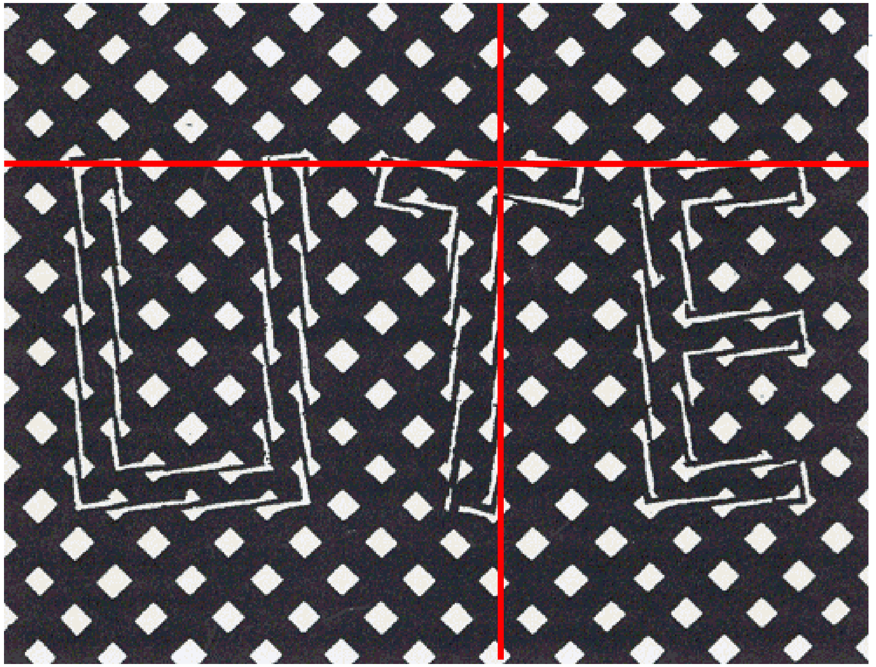
High(er) level vision

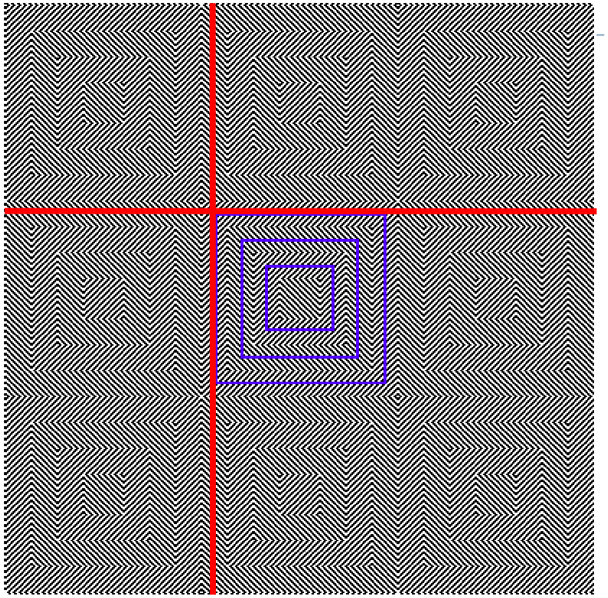
Simultaneous contrast

High-Level Contrast Processing

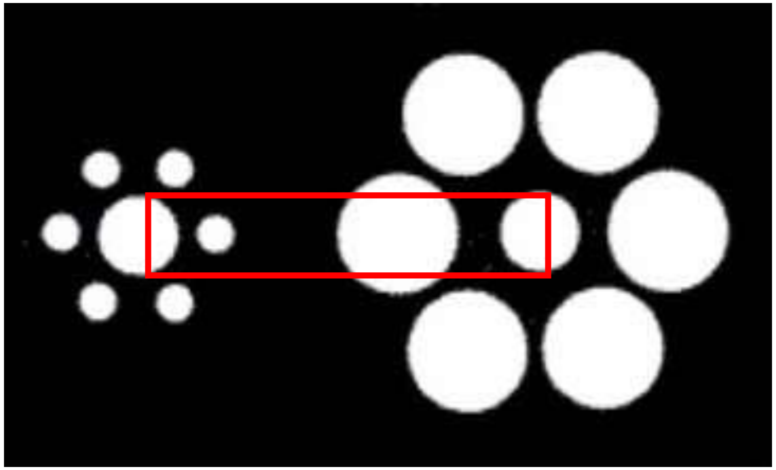


Shape Perception



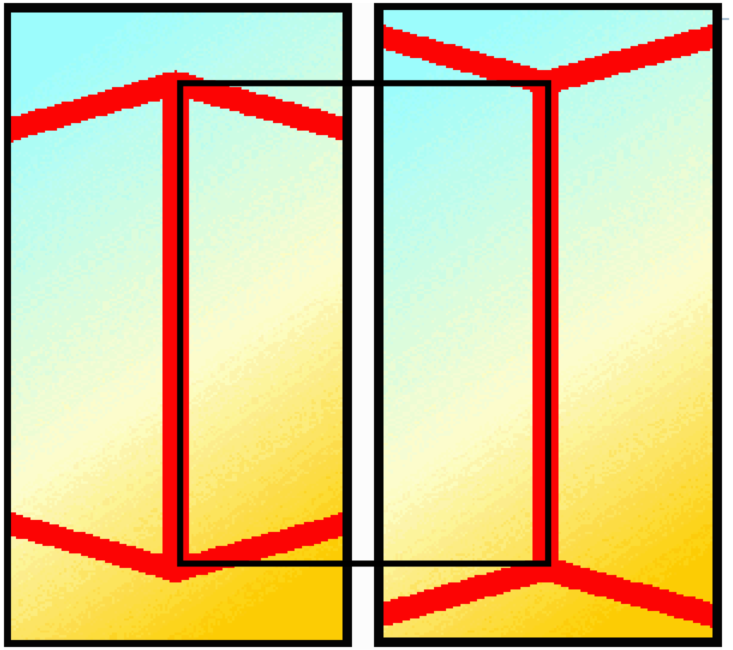


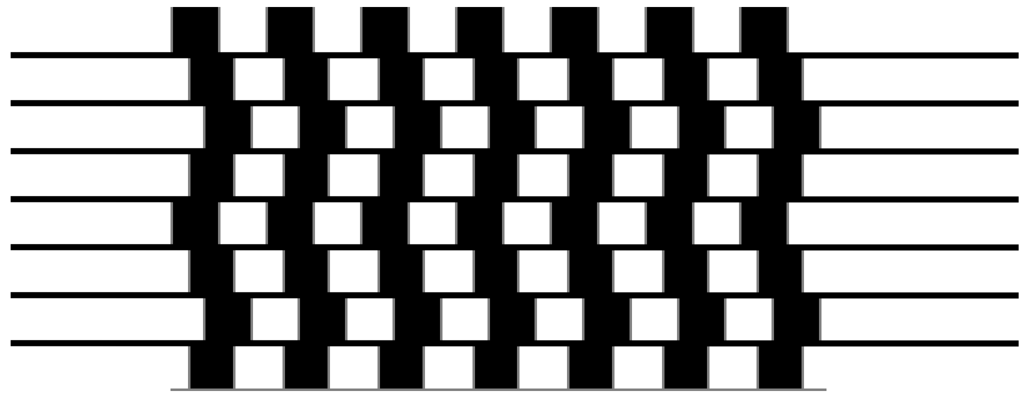
- Depends on surrounding primitives
  - Directional emphasis
  - Size emphasis



http://www.panoptikum.net/optischetaeuschungen/index.html

Shape Processing: Geometrical Clues



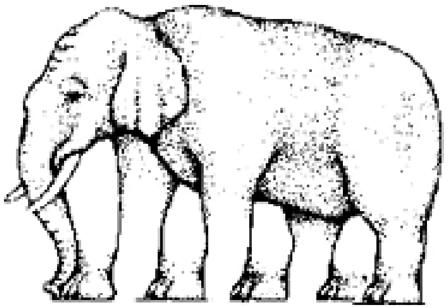


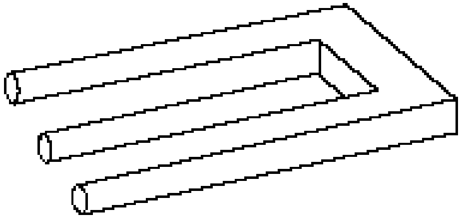
http://www.panoptikum.net/optischetaeuschungen/index.html

- Automatic geometrical interpretation
  - 3D perspective
  - Implicit scene depth

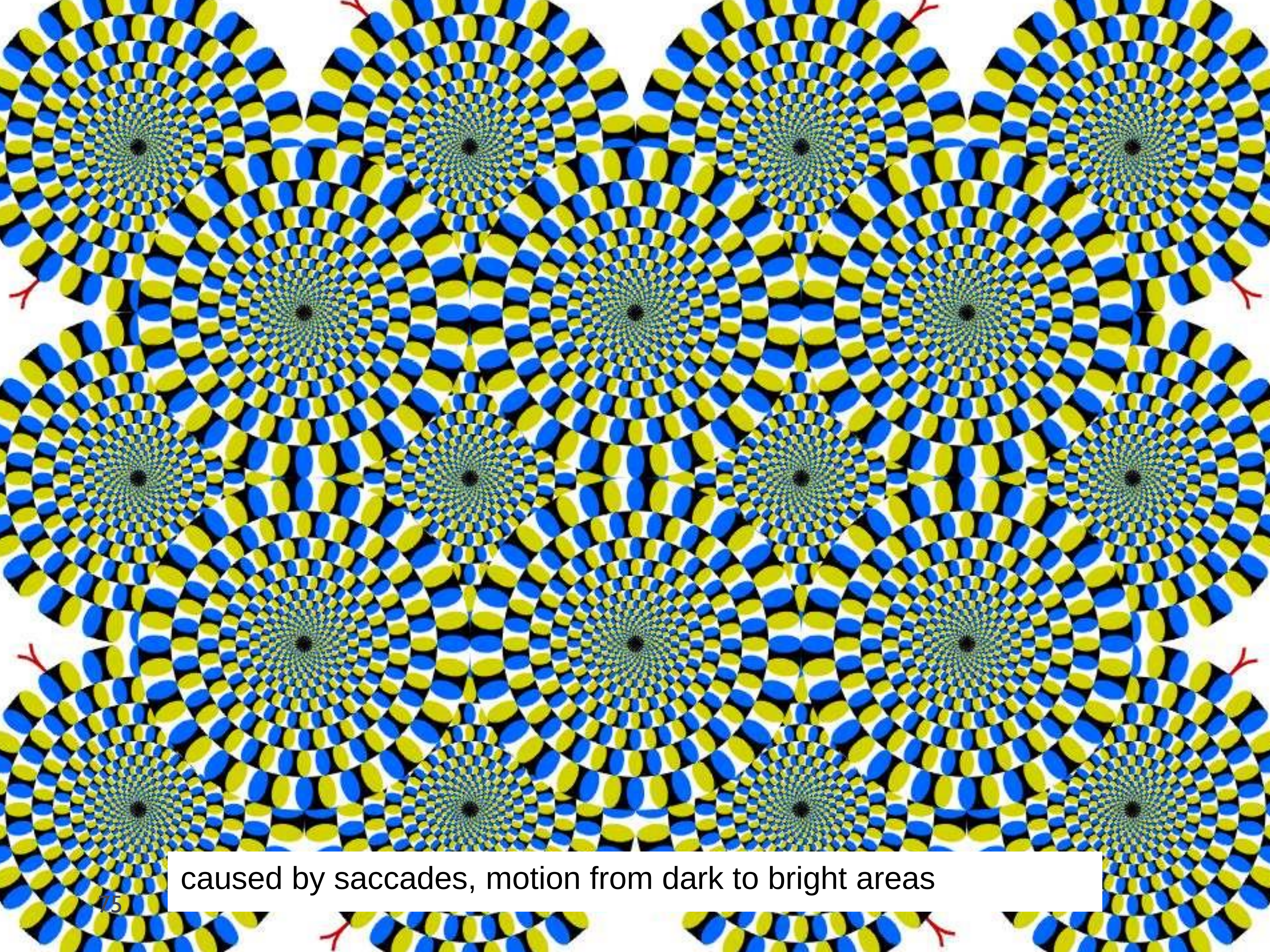
Impossible Scenes

- Escher et.al.
  - Confuse HVS by presenting contradicting visual clues
  - Local vs. global processing





http://www.panoptikum.net/optischetaeuschungen/index.html



caused by saccades, motion from dark to bright areas

Law of closure

References

- Wandell, B. A. (1995). *Foundations of vision*. Sinauer Associates.
- Mantiuk, R. K., Myszkowski, K., & Seidel, H. (2015). High Dynamic Range Imaging. In *Wiley Encyclopedia of Electrical and Electronics Engineering*. Wiley.
  - Section 2.4
  - Available online: http://www.cl.cam.ac.uk/~rkm38/hdri_book.html