Models of early visual perception

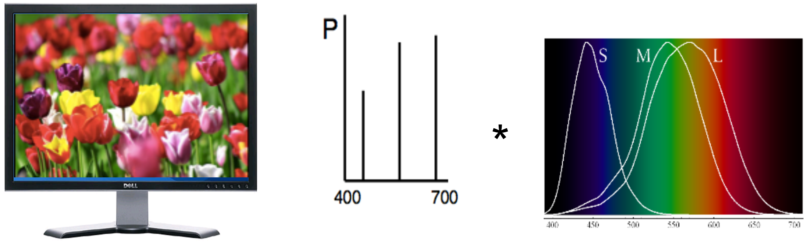
Advanced Graphics

Rafal Mantiuk
Computer Laboratory, University of Cambridge

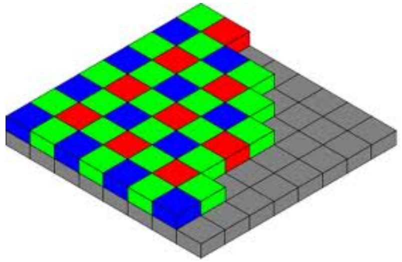
Many technical solutions are motivated by visual perception

H.264
MPEG-4/AVC

Image & video compression

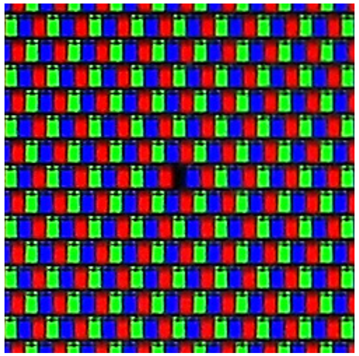


Display spectral emission - metamerism

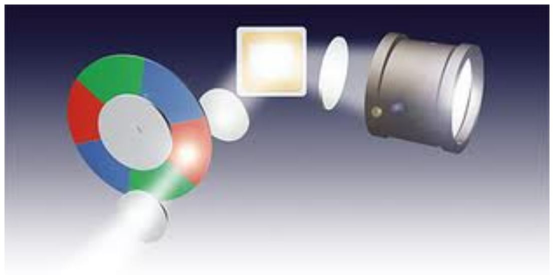


Camera's Bayer pattern

Halftonning



Display's subpixels



Color wheel in DLPs

…

Perceived brightness of light

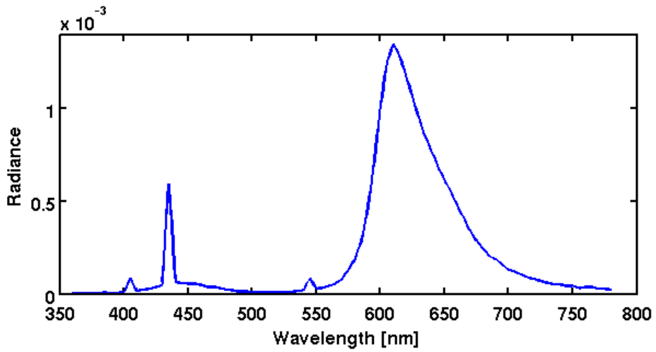
Luminance

- Luminance – how bright the surface will appear regardless of its colour. Units: cd/m^2

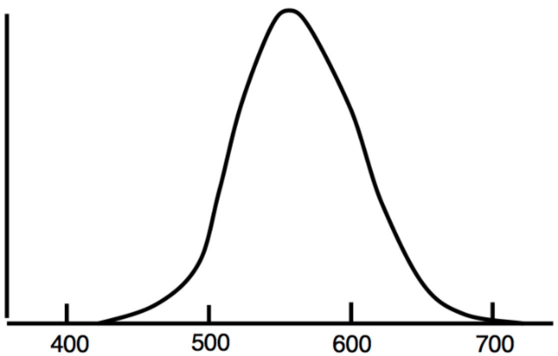
Luminance

$$L_V = \int_0^{\infty} L(\lambda) \cdot V(\lambda) d\lambda$$

Light spectrum (radiance)



Luminous efficiency function (weighting)



Luminance and Luma

- Luminance
  - Photometric quantity defined by the spectral luminous efficiency function
- L ≈ 0.2126 R + 0.7152 G + 0.0722 B
- Units: cd/m^2

- Luma
  - Gray-scale value computed from LDR (gamma corrected) image
- Y = 0.2126 R' + 0.7152 G' + 0.0722 B'
  - R' – prime denotes gamma correction

$$R' = R^{1/\gamma}$$

- Unitless

Gamma correction (reminder – Color lecture)

- Gamma correction is used to encode luminance or tri-stimulus color values (RGB) in imaging systems (displays, printers, cameras, etc.)

Gain

Gamma (usually =2.2)

$$V_{out} = a \cdot V_{in}^{\gamma}$$

Luminance

Luma

For color images: $R = a \cdot (R')^{\gamma}$ and the same for green and blue

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:

Perceived magnitude

$$\varphi(I) = kI^{a}$$

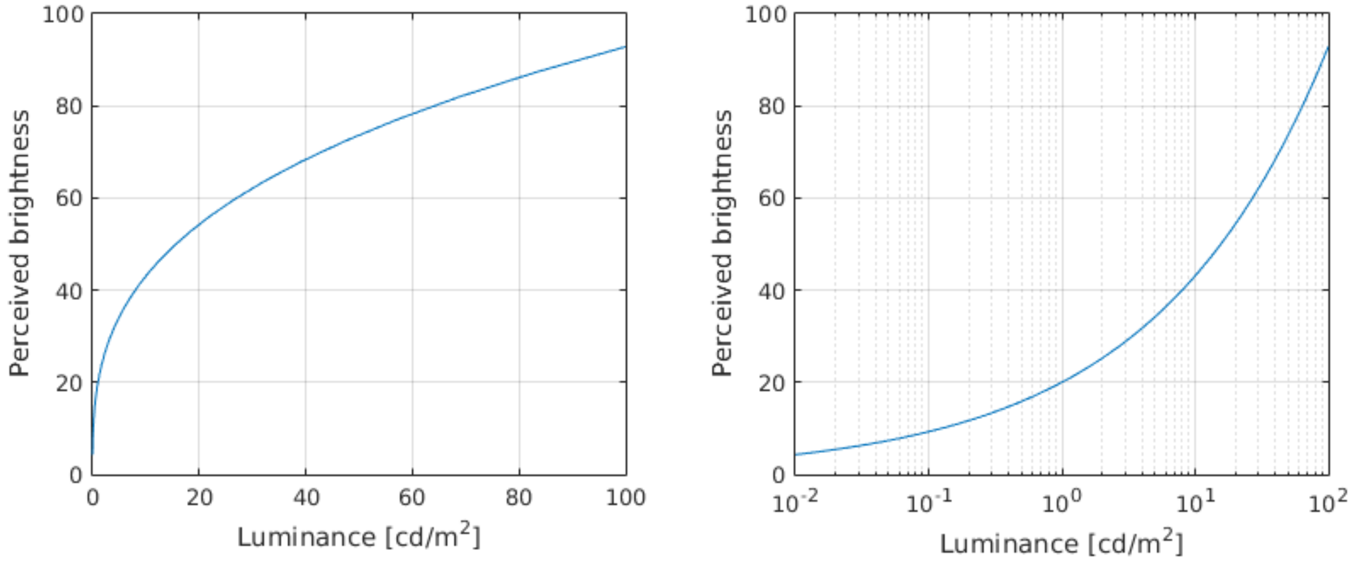
Exponent

Constant

Physical stimulus

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

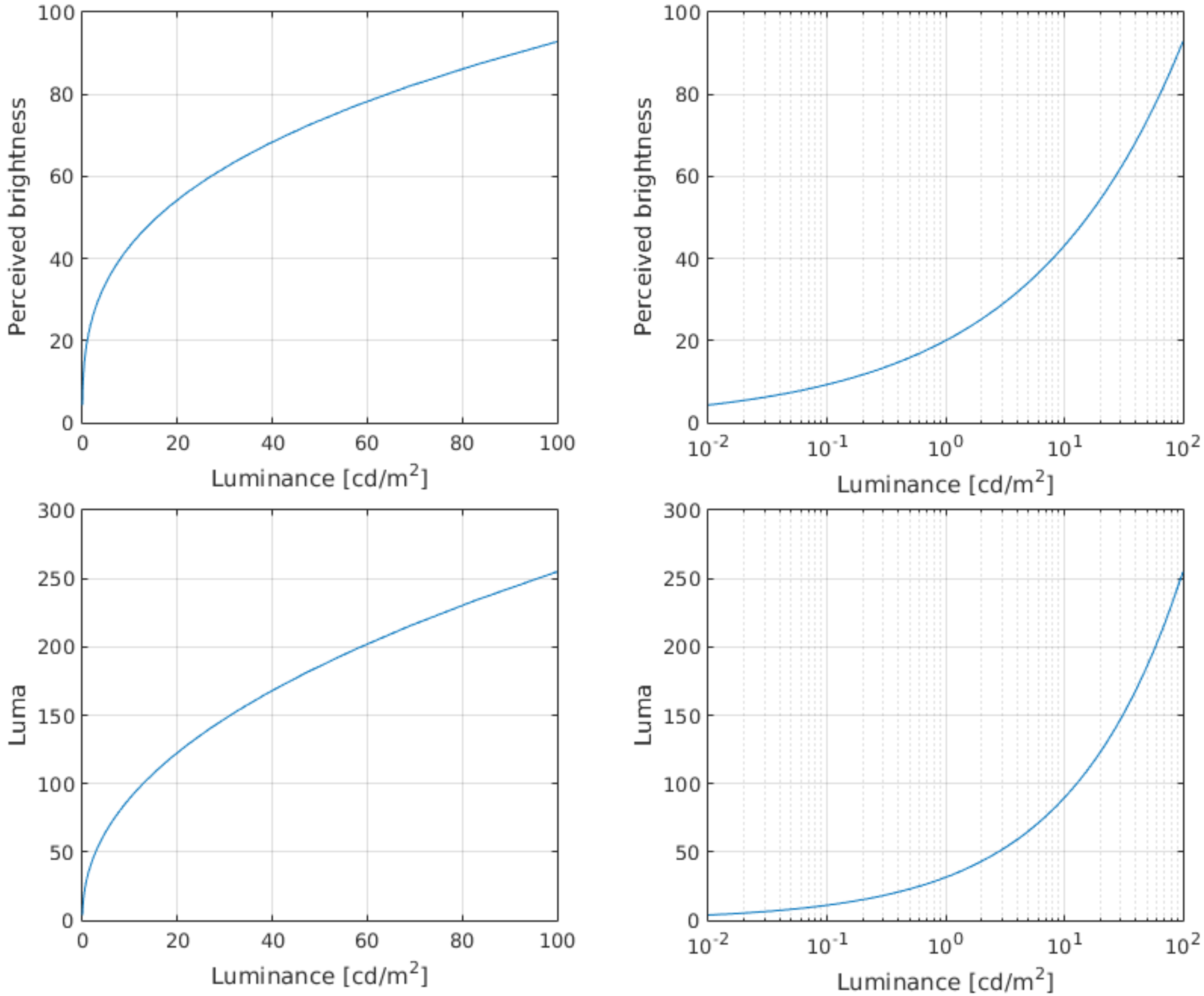
Steven’s law for brightness



Steven’s law vs. Gamma correction

Stevens’ law
a=0.3

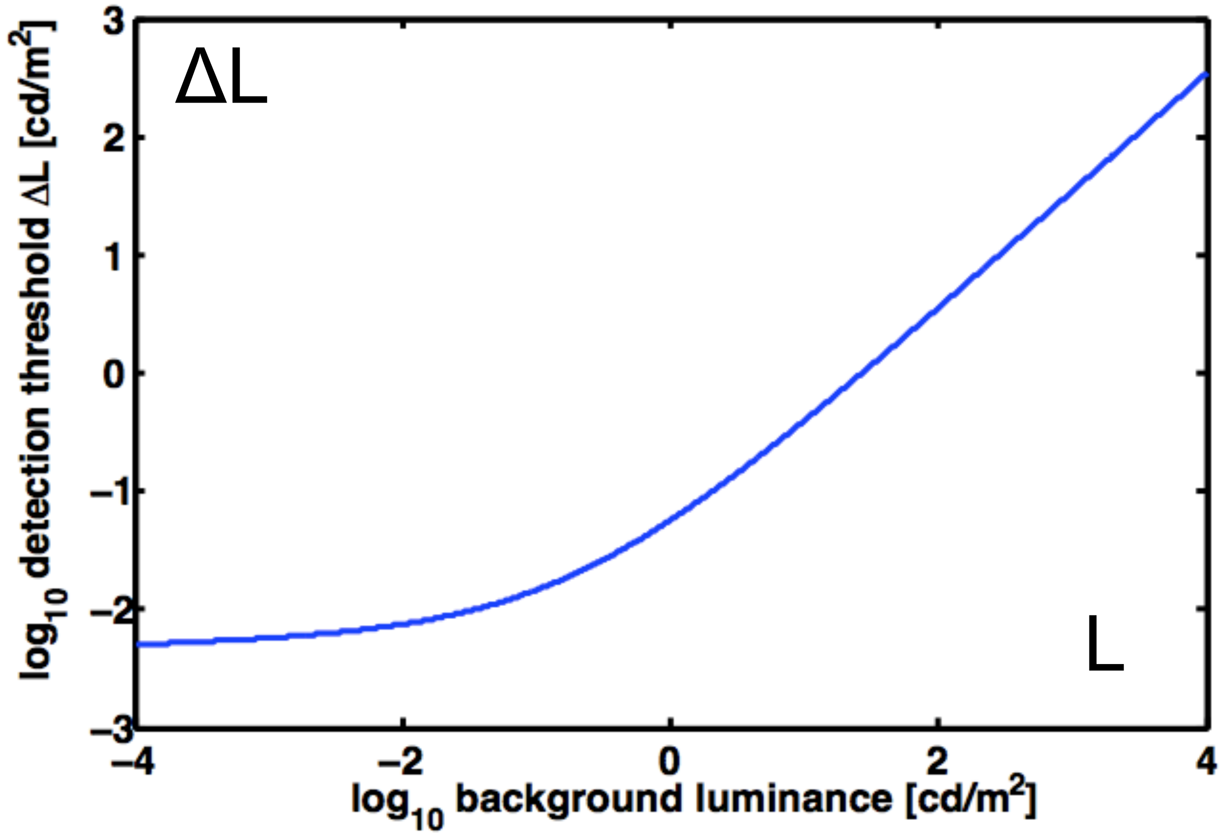
Gamma function
Gamma = 2.2

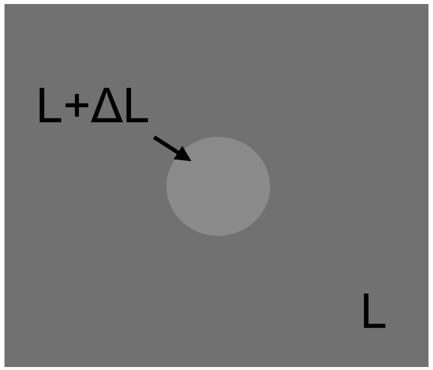


Detection and discrimination

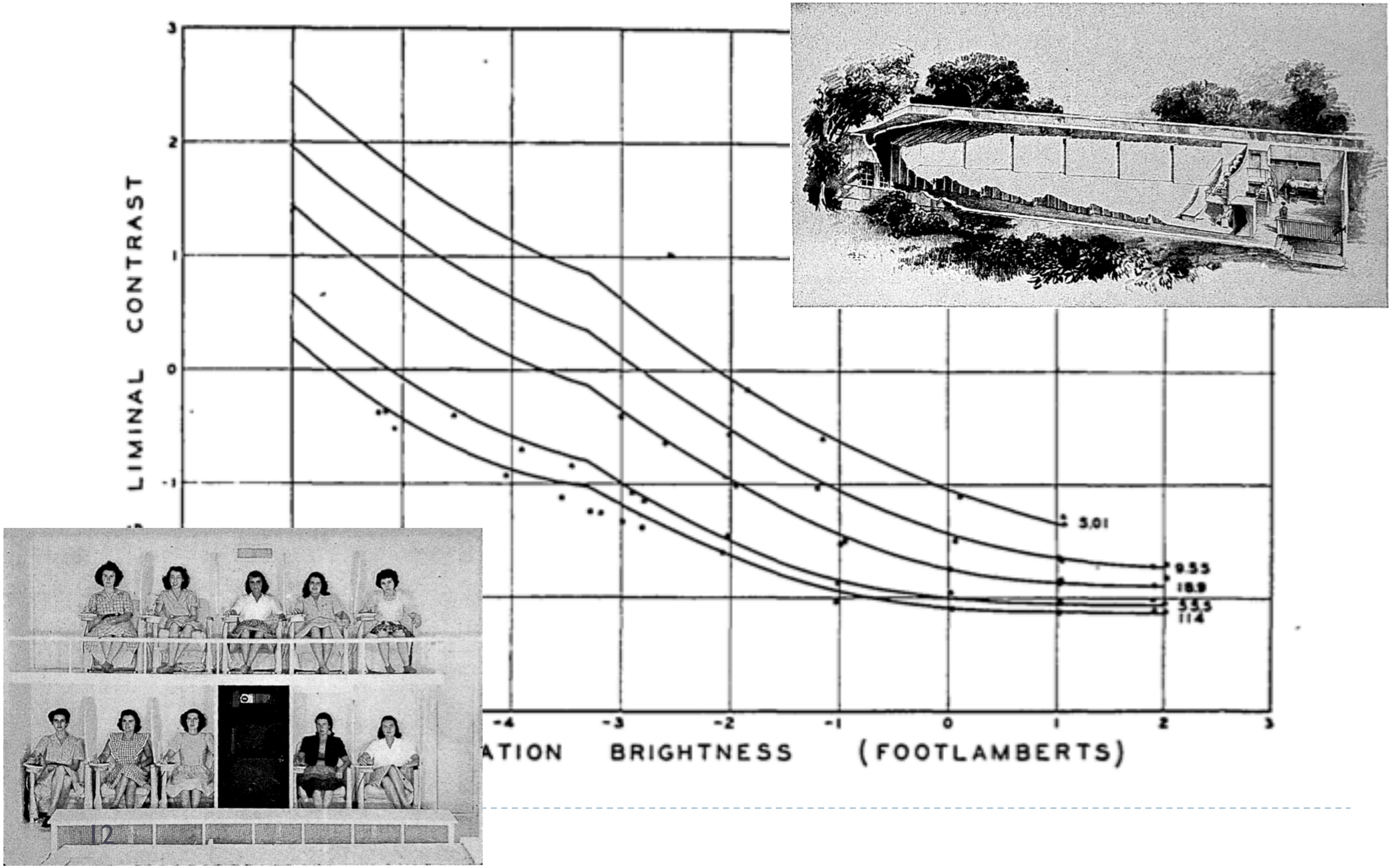
Threshold versus intensity (t.v.i.) function

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

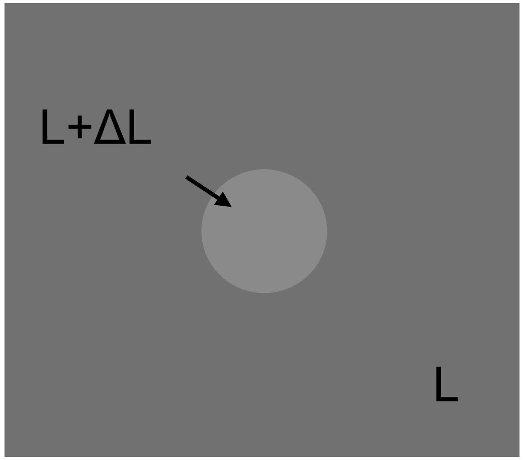


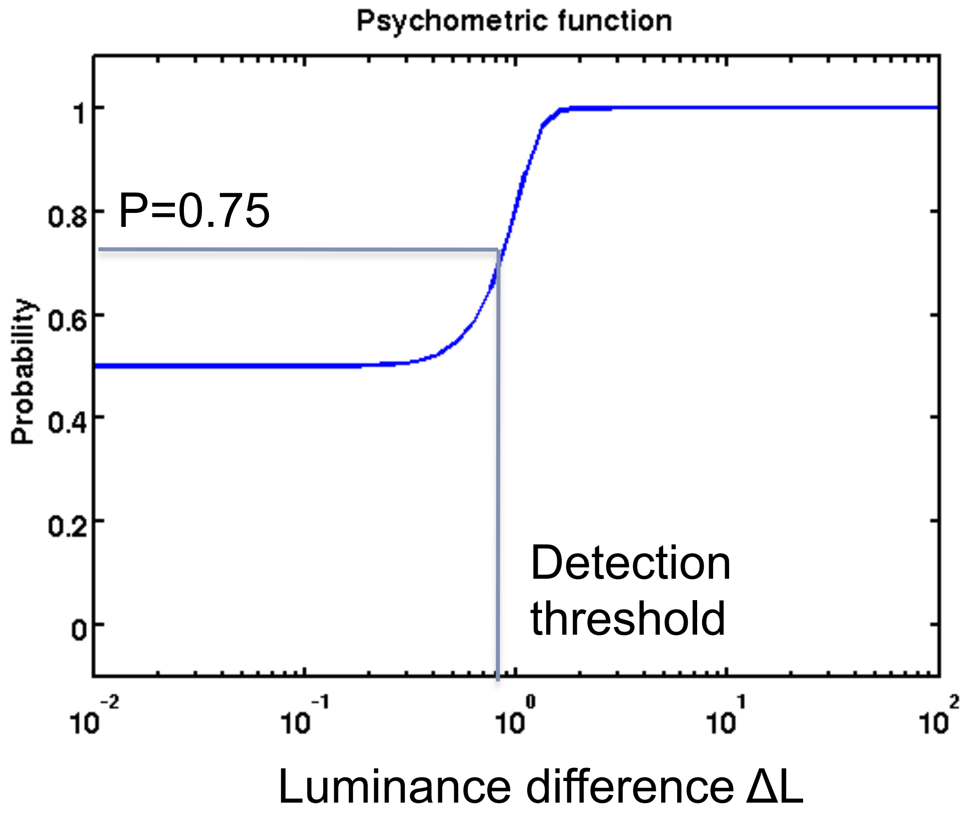


t.v.i. measurements – Blackwell 1946



Psychophysics
Threshold experiments

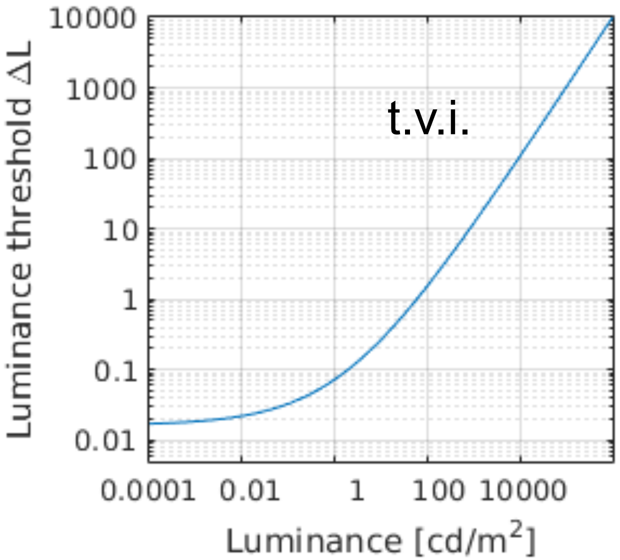




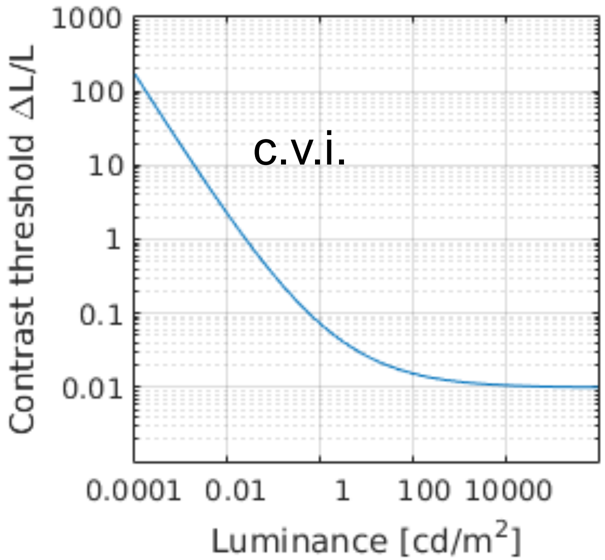
t.v.i function / c.v.i. function / Sensitivity

- The same data, different representation

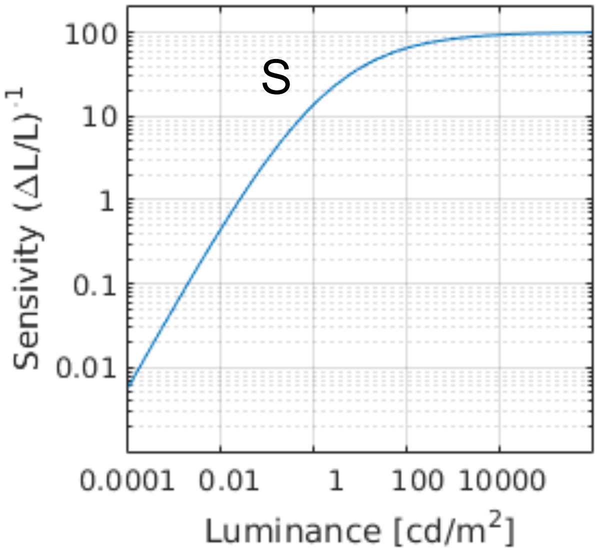
Threshold vs. intensity



Contrast vs. intensity



Sensitivity



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

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

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

Sensitivity to luminance

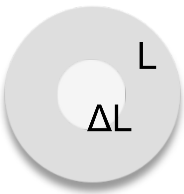
- Weber-law – the just-noticeable difference is proportional to the magnitude of a stimulus

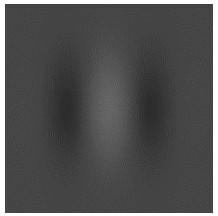
Ernst Heinrich Weber
[From wikipedia]

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

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

Typical stimuli:





Consequence of the Weber-law

- Smallest detectable difference in luminance

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

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!

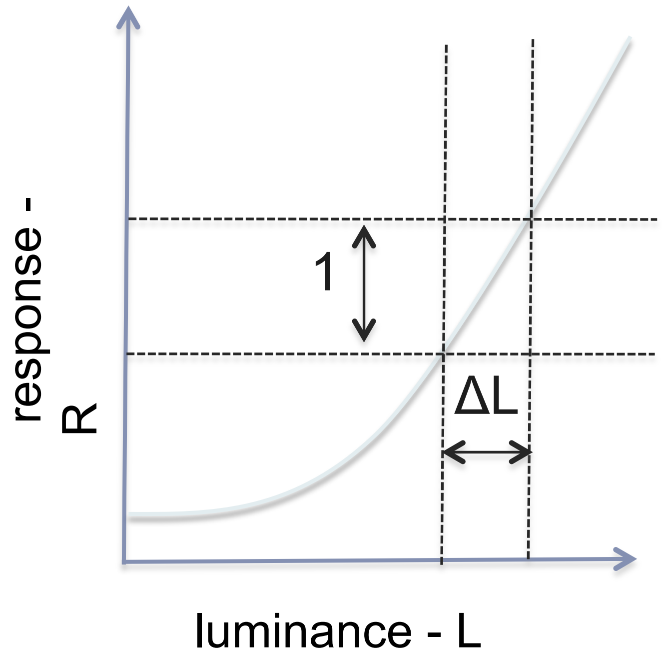
How to make luminance (more) perceptually uniform?

- Using "Fechnerian" integration

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

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

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

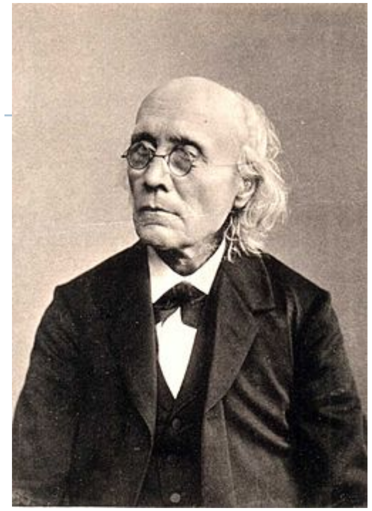
- the response of the visual system to light is:

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

Fechner law

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

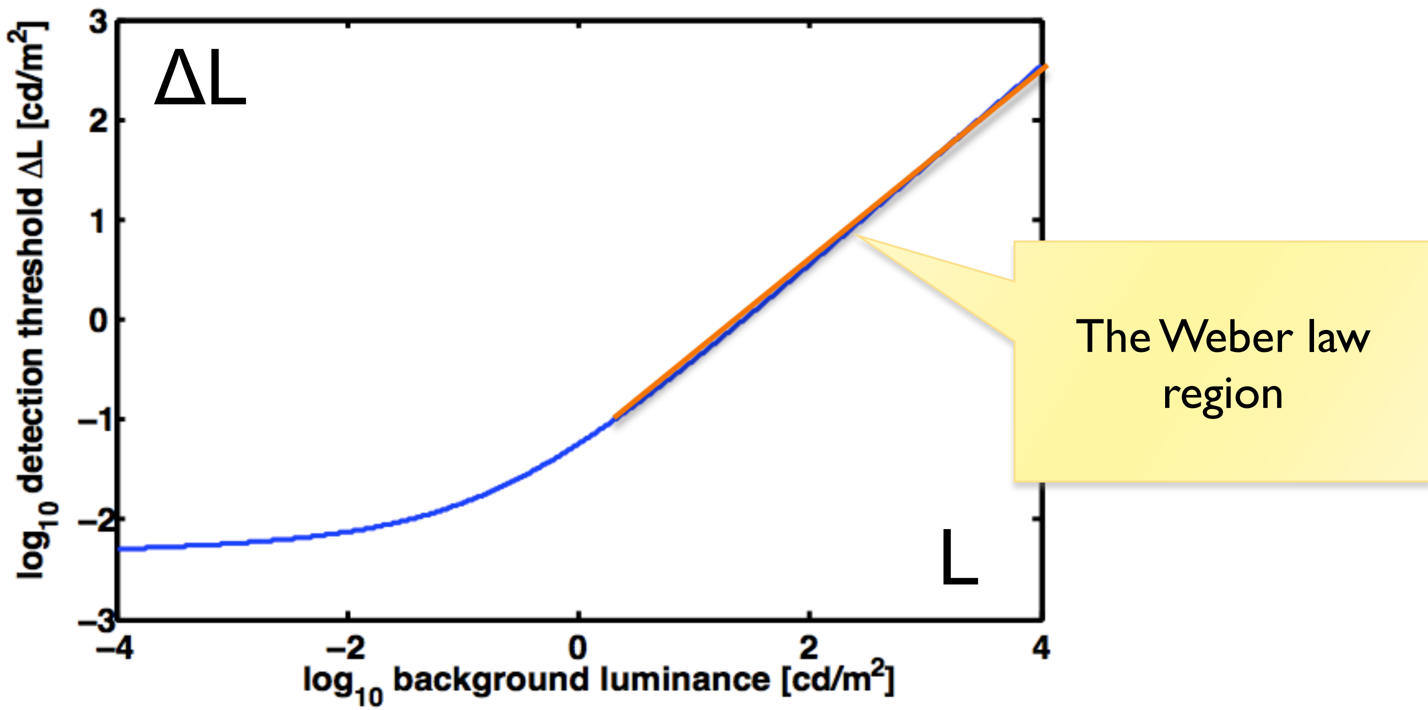
- Response of the visual system to luminance is **approximately** logarithmic



Gustav Fechner
[From Wikipedia]

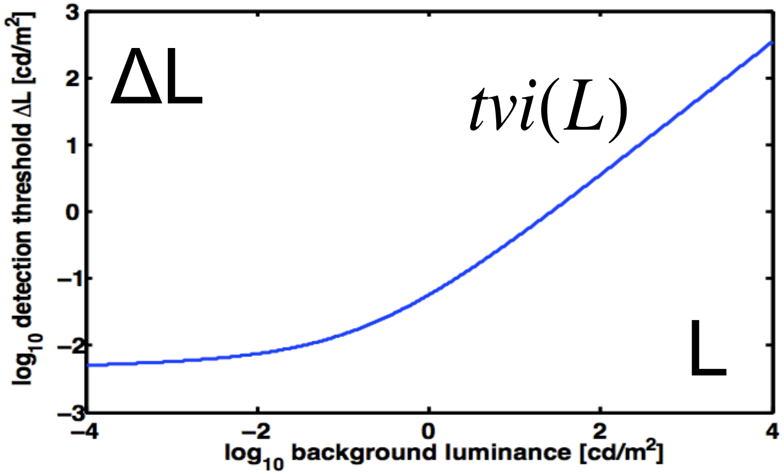
But…the Fechner law does not hold for the full luminance range

- 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:

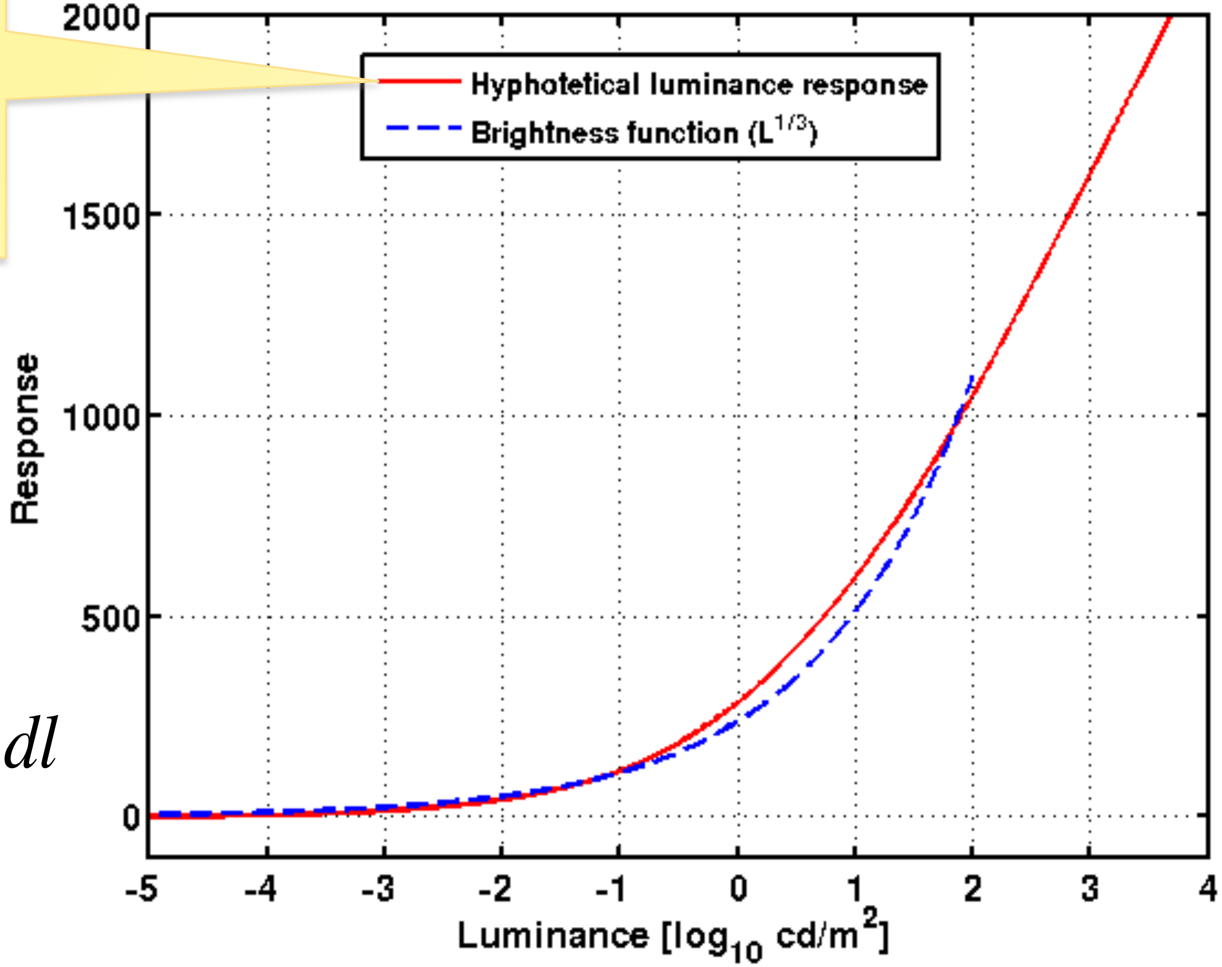


- we can get more accurate estimate of the "response":

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

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)

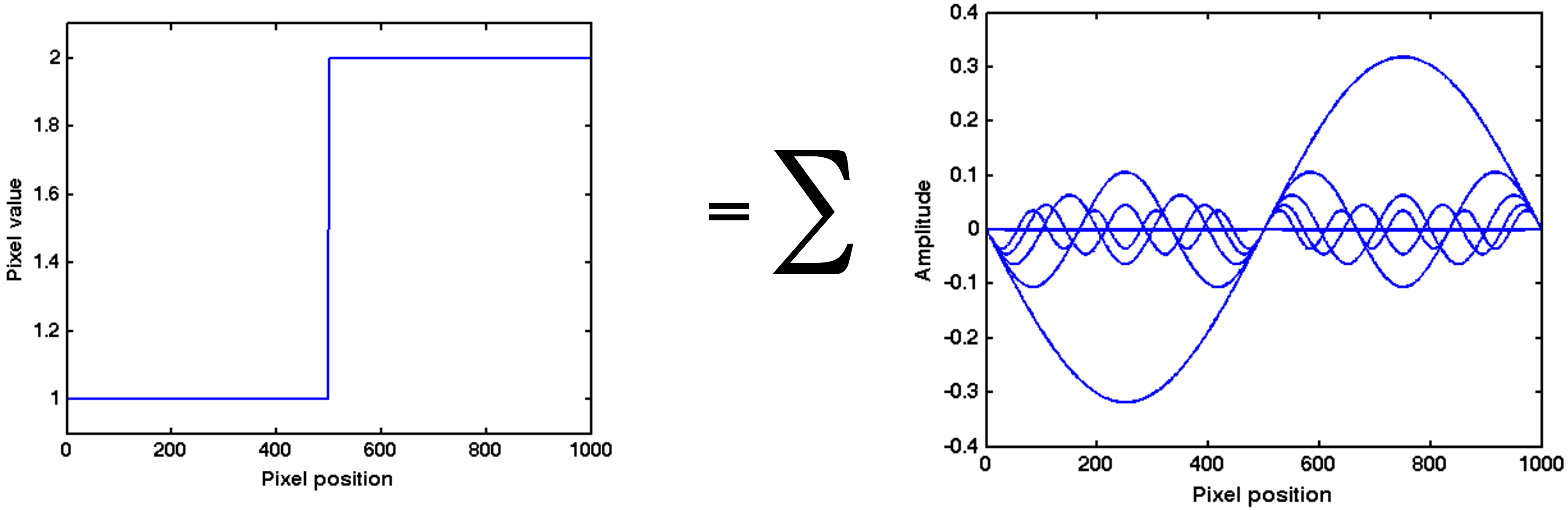
- DICOM grayscale function
  - Function used to encode signal for medial monitors
  - 10-bit JND-scaled (just noticeable difference)
  - Equal visibility of gray levels
- Dolby Vision – Perceptual Quantizer
  - To encode pixels for high dynamic range images and video

DOLBY VISION
The Future of Vision

Spatial contrast sensitivity

Fourier analysis

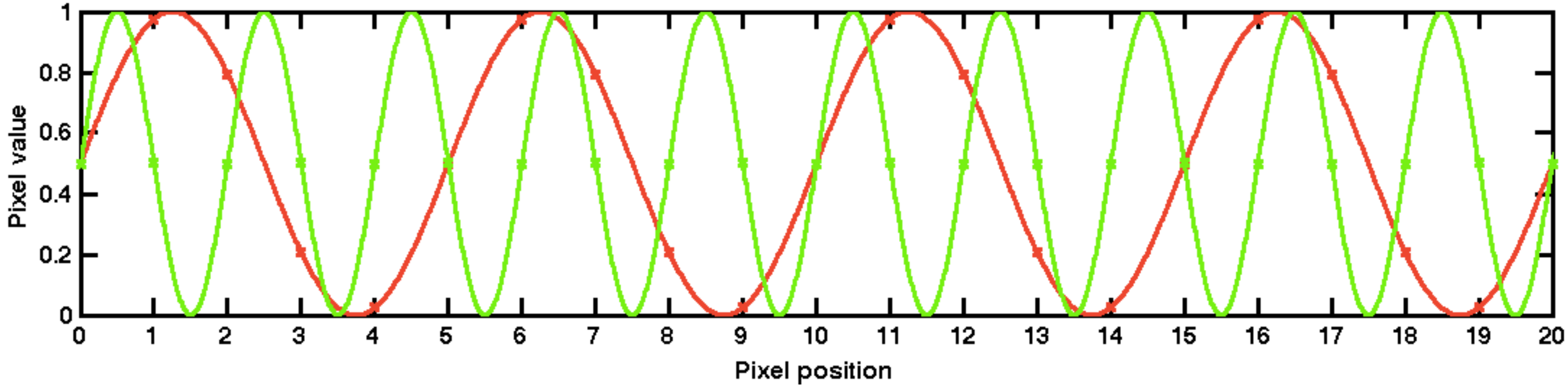
- Every N-dimensional function (including images) can be represented as a sum of sinusoidal waves of different frequency and phase



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

Spatial frequency in images

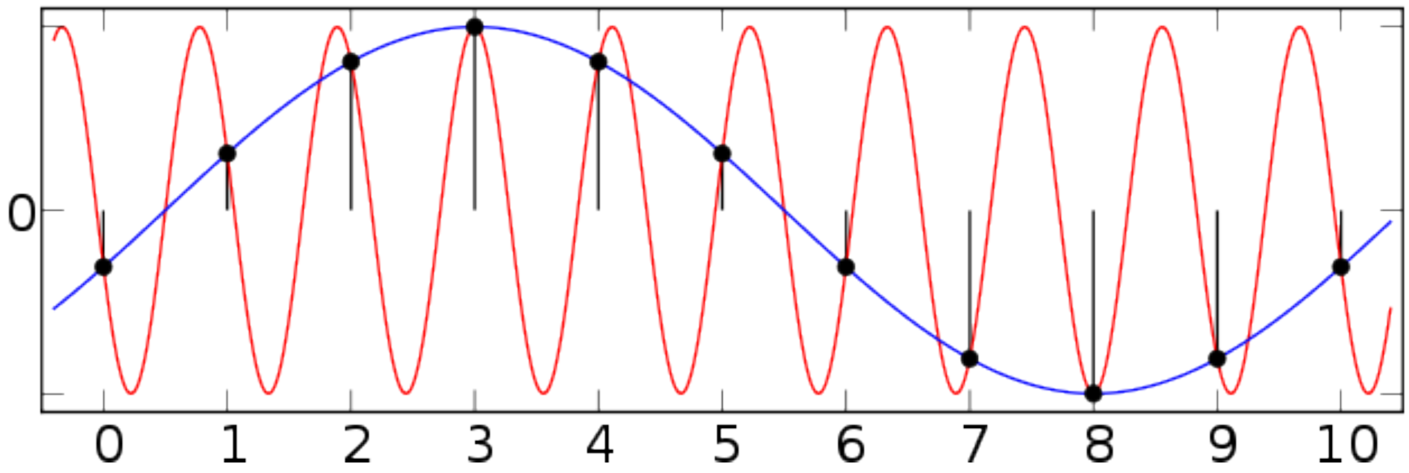
- Image space units: cycles per sample (or cycles per pixel)



  - 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

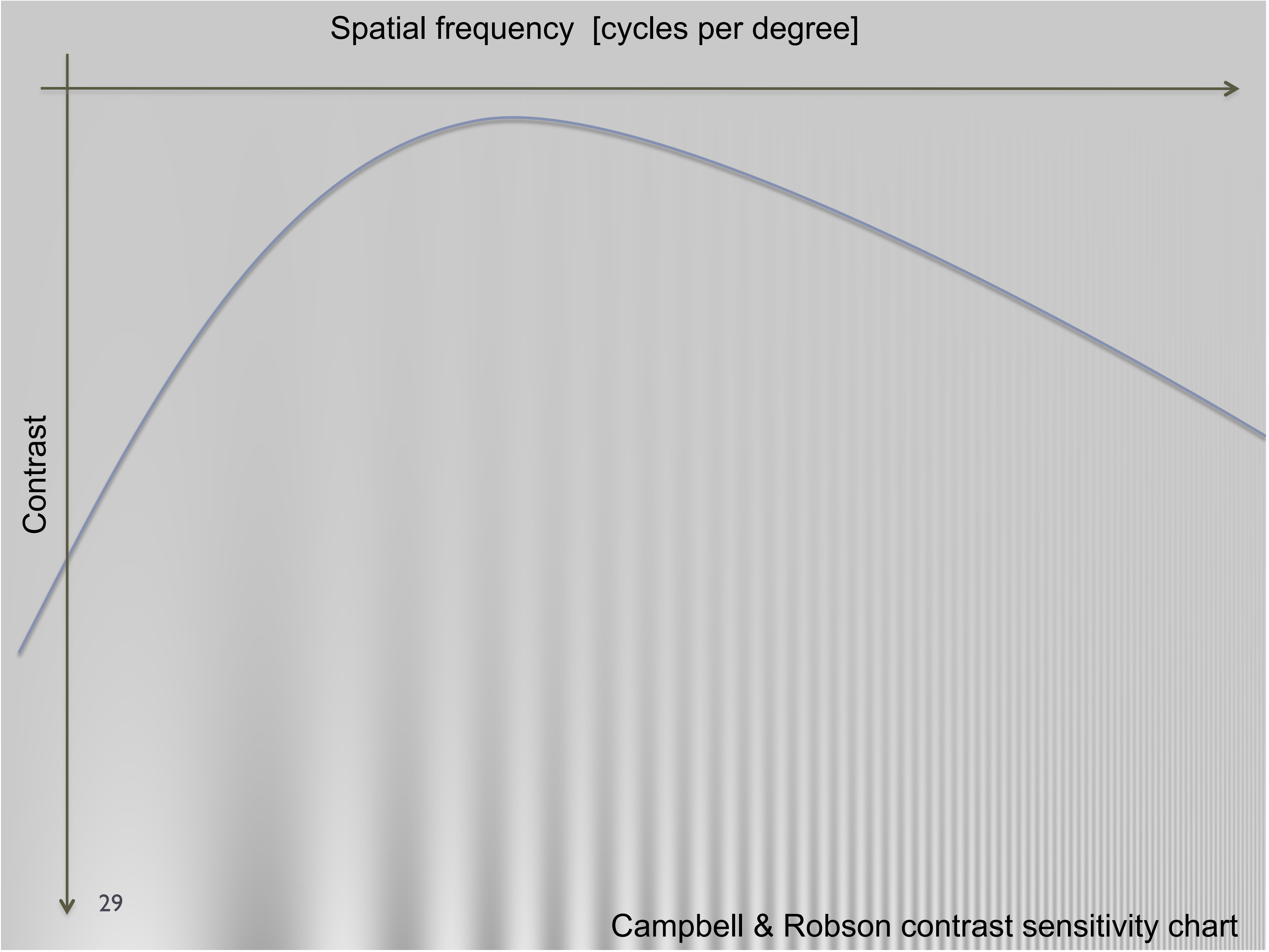
- What is the highest frequency that can be reconstructed for a given sampling density?
  - Sampling density – how many pixels per image/visual angle/…

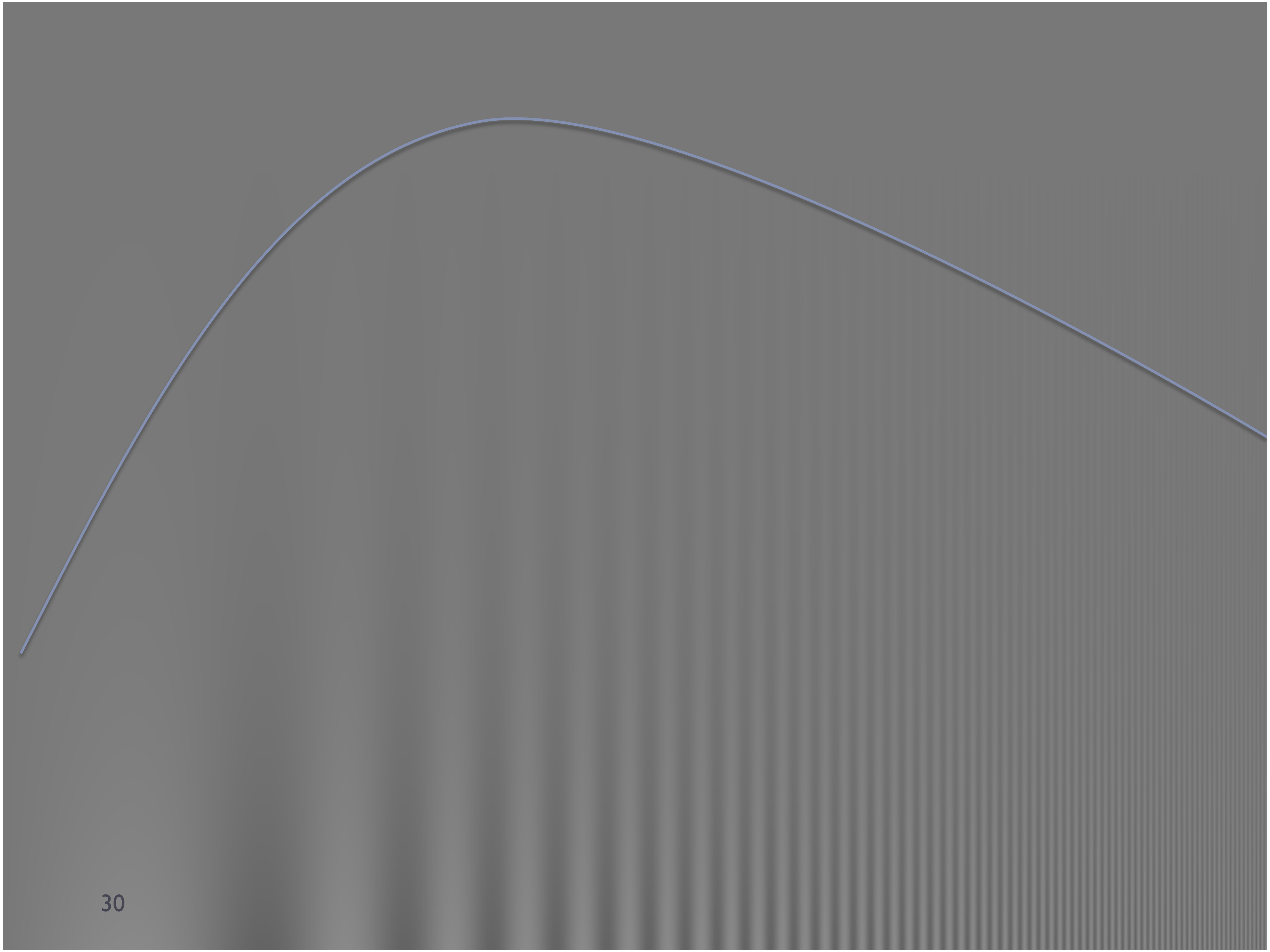


  - Two different sinusoids can be fitted to this set of samples
  - It is possible to fit an infinite number of sinusoids if we allow infinitely high frequency

Nyquist frequency / aliasing

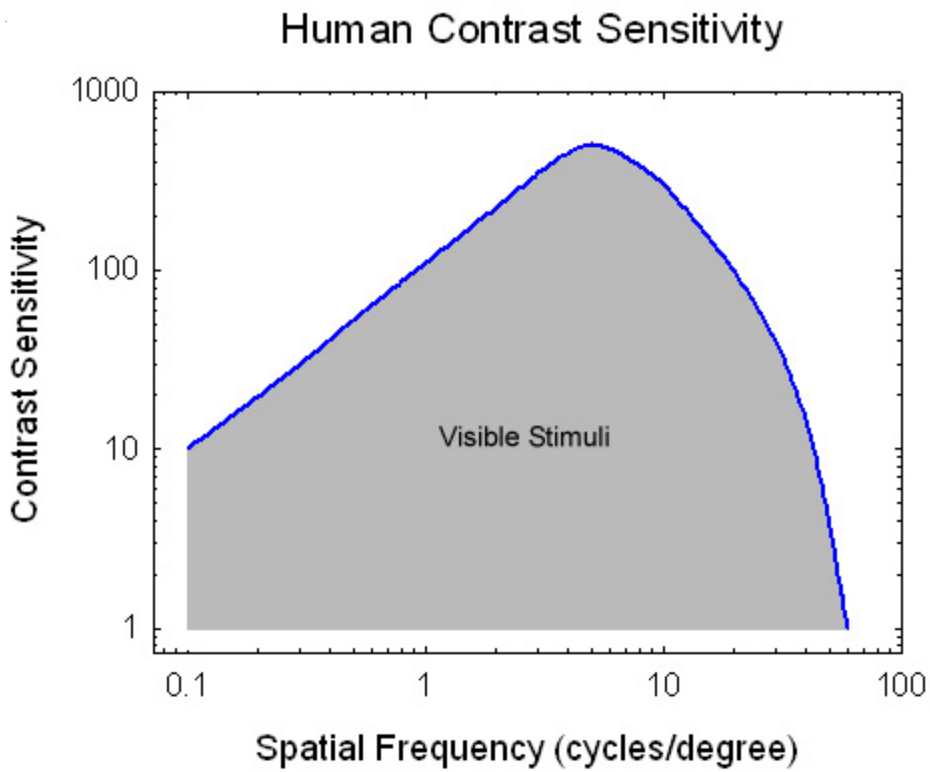
- Nuquist frequency is the highest frequency that can be represented by a discrete set of uniform samples (pixels)
- Nuquist frequency = 0.5 sampling rate
  - For audio
    - If the sampling rate is 44100 samples per second (audio CD), then the Nyquist frequency is 22050 Hz
  - For images (visual degrees)
    - 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


Spatial frequency [cycles per degree]
Contrast
Campbell & Robson contrast sensitivity chart



Contrast Sensitivity

- Sensitivity:
  1 / threshold contrast
- Contrast = Δ L/L
- Maximum acuity 2-5 cycles/degree (0.2 %)
  - Decrease toward low frequencies: lateral inhibition
  - Decrease toward high frequencies
  - Upper limit: 60-70 cycles/ degree



Rationale: if we were sensitive to low frequencies, the vision would be affected by changes of illumination. There are physical limitations to the perception of high frequencies.

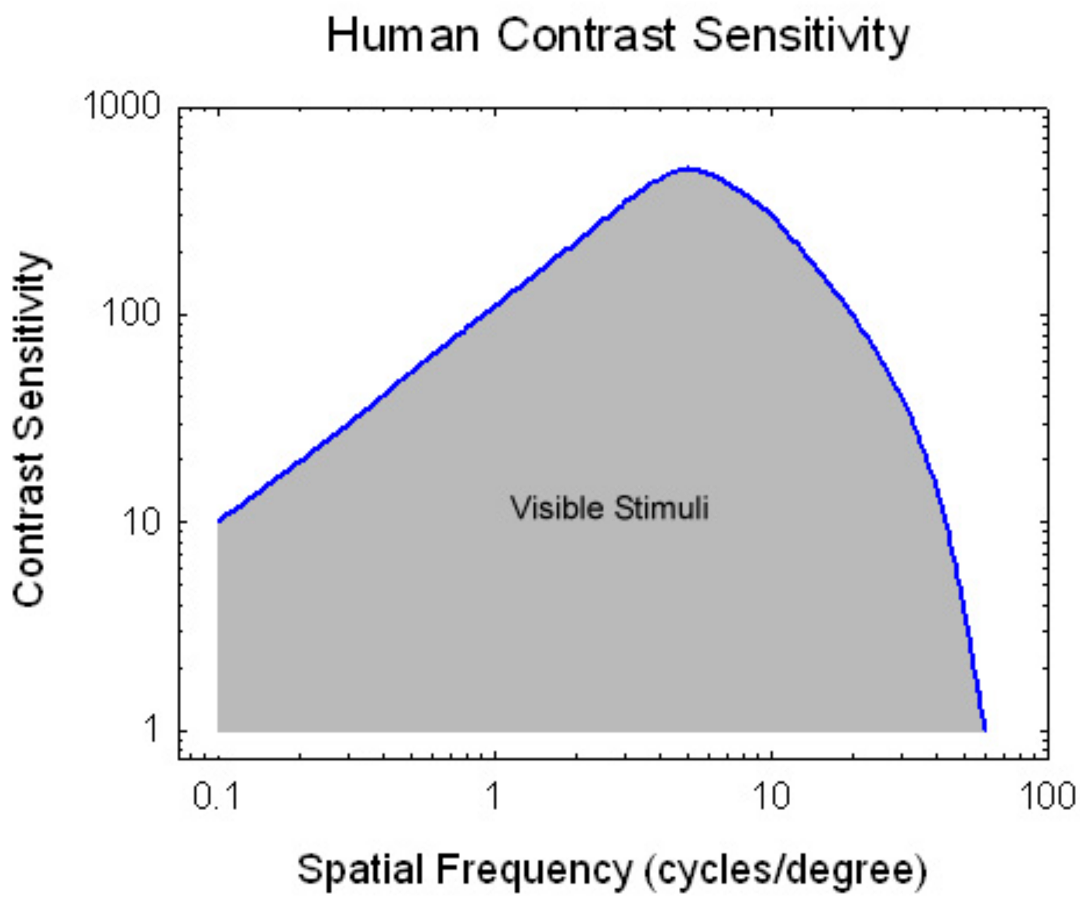
www.psychology.psych.ndsu.nodak.edu

Implications of CSF

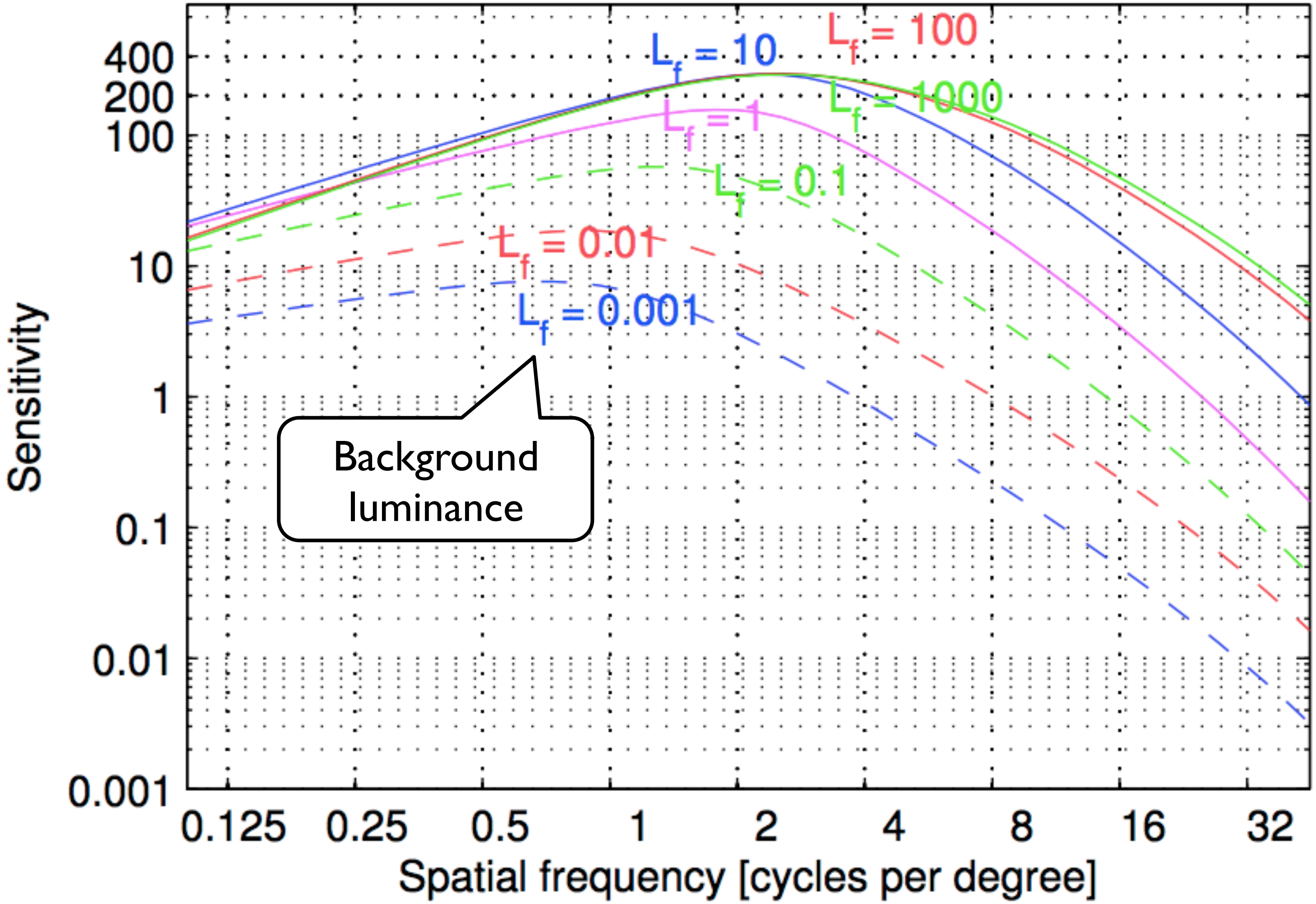
- As objects get further away, they get smaller; spatial frequencies get higher
  - At some point we cannot see the details
  - That is the upper limit of the CSF (60-70 cpd)
- When we get to close to low frequency patterns, they seem to be constant
  - The background of this slide contains a smooth gradient
  - It is well visible when you look at it from a normal viewing distance
  - Now enlarge the slide to full screen and move your head very close
  - The gradient should disappear

Explaining the effects with CSF

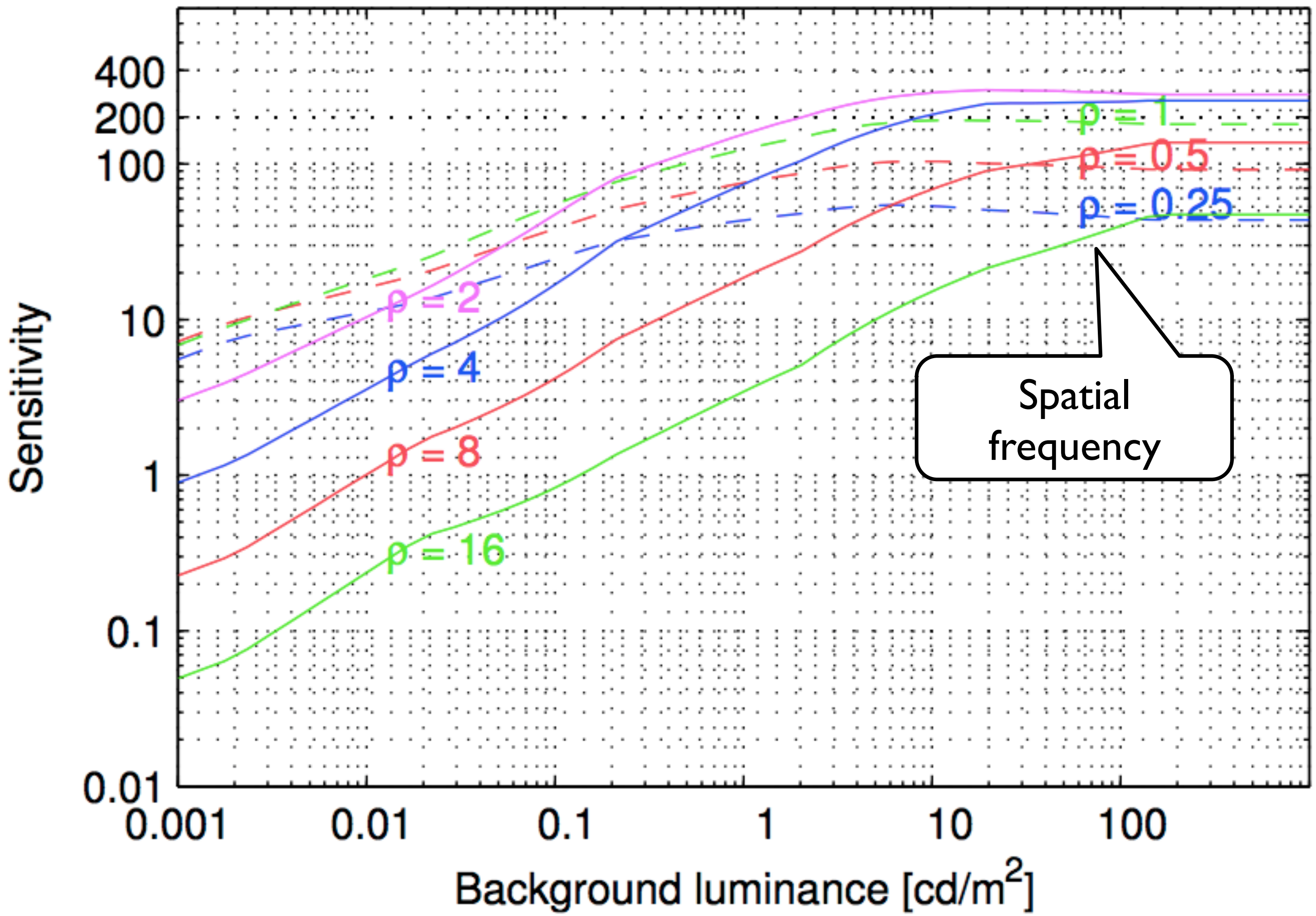
- Can you explain the effects described on the previous slide using the CSF plot?



CSF as a function of spatial frequency



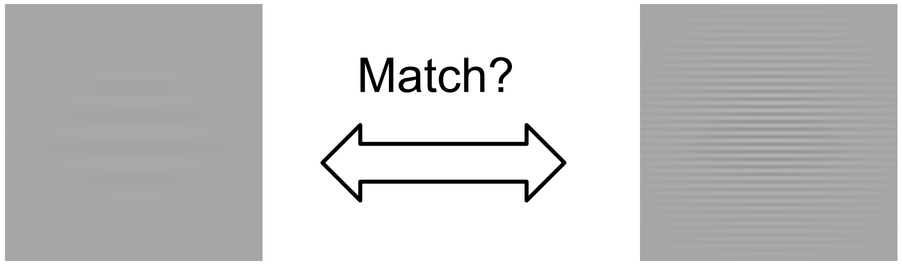
CSF as a function of background luminance

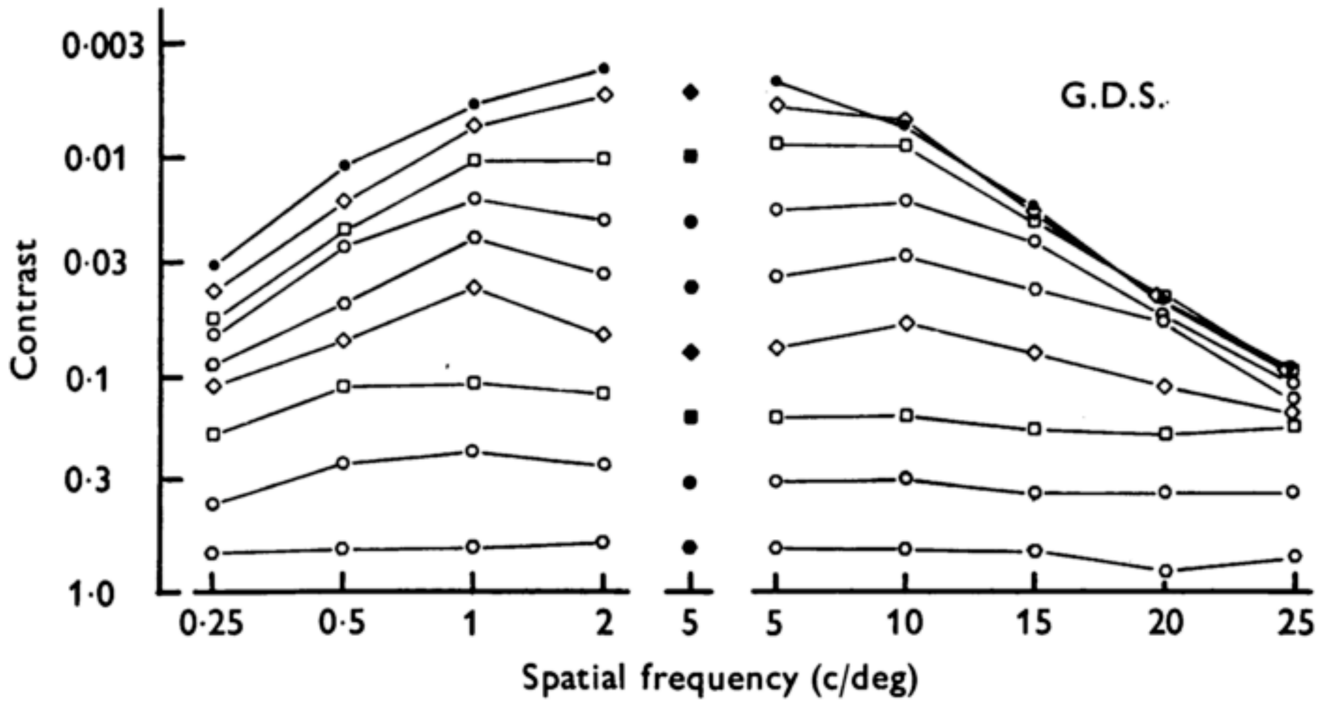


Contrast constancy

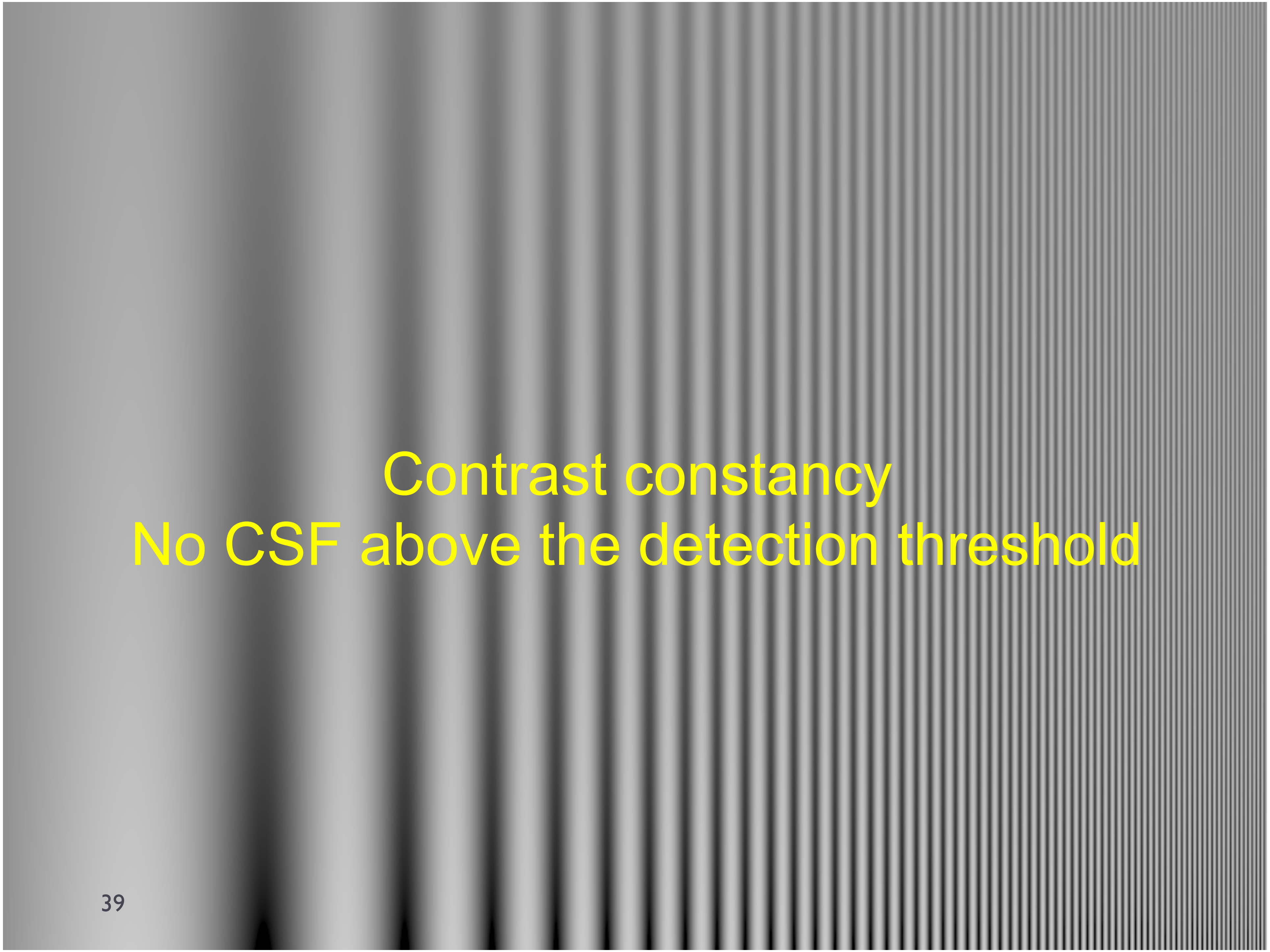
Contrast constancy

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



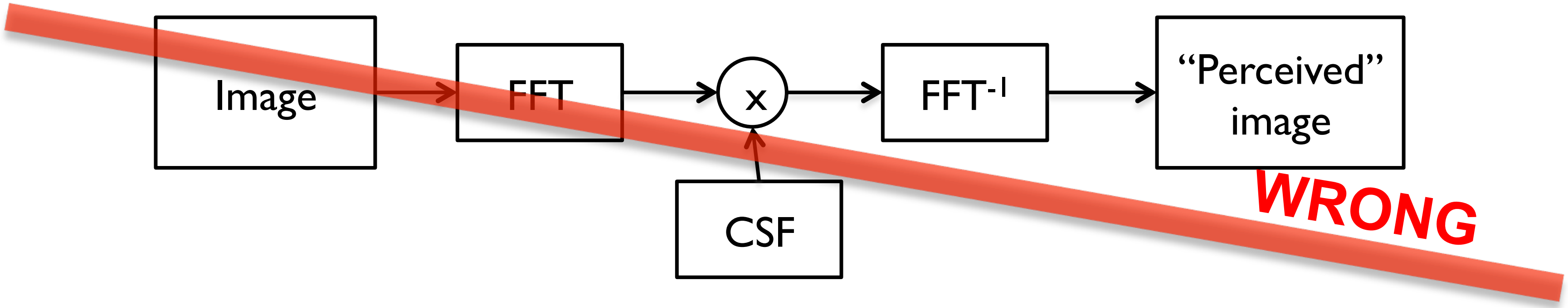


From: Georgeson and Sullivan. 1975. J. Phsysio.


Contrast constancy
No CSF above the detection threshold

Contrast constancy

- For high (supra-threshold) contrast, the perceived magnitude of contrast does not change with spatial frequency
- The CSF gets "flat" above the detection threshold
- A common mistake - use CSF as a linear filter:

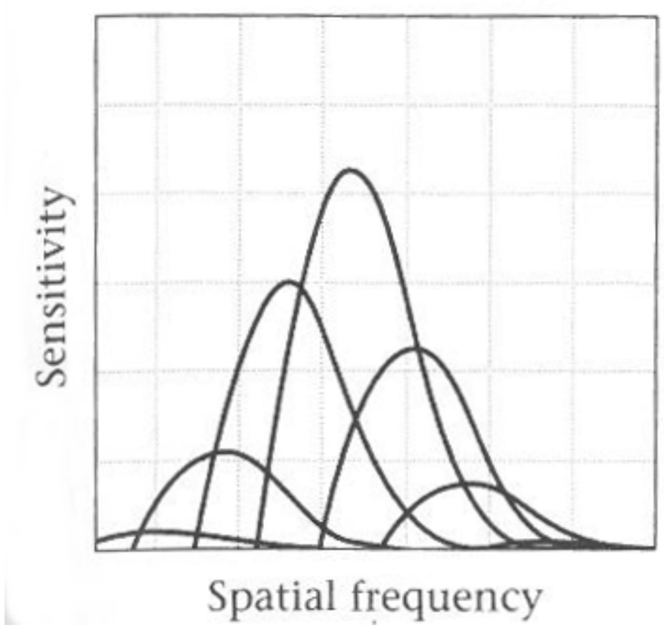


  - why CSF cannot be used as a linear filter?

Multi-resolution models

Spatial-frequency selective channels

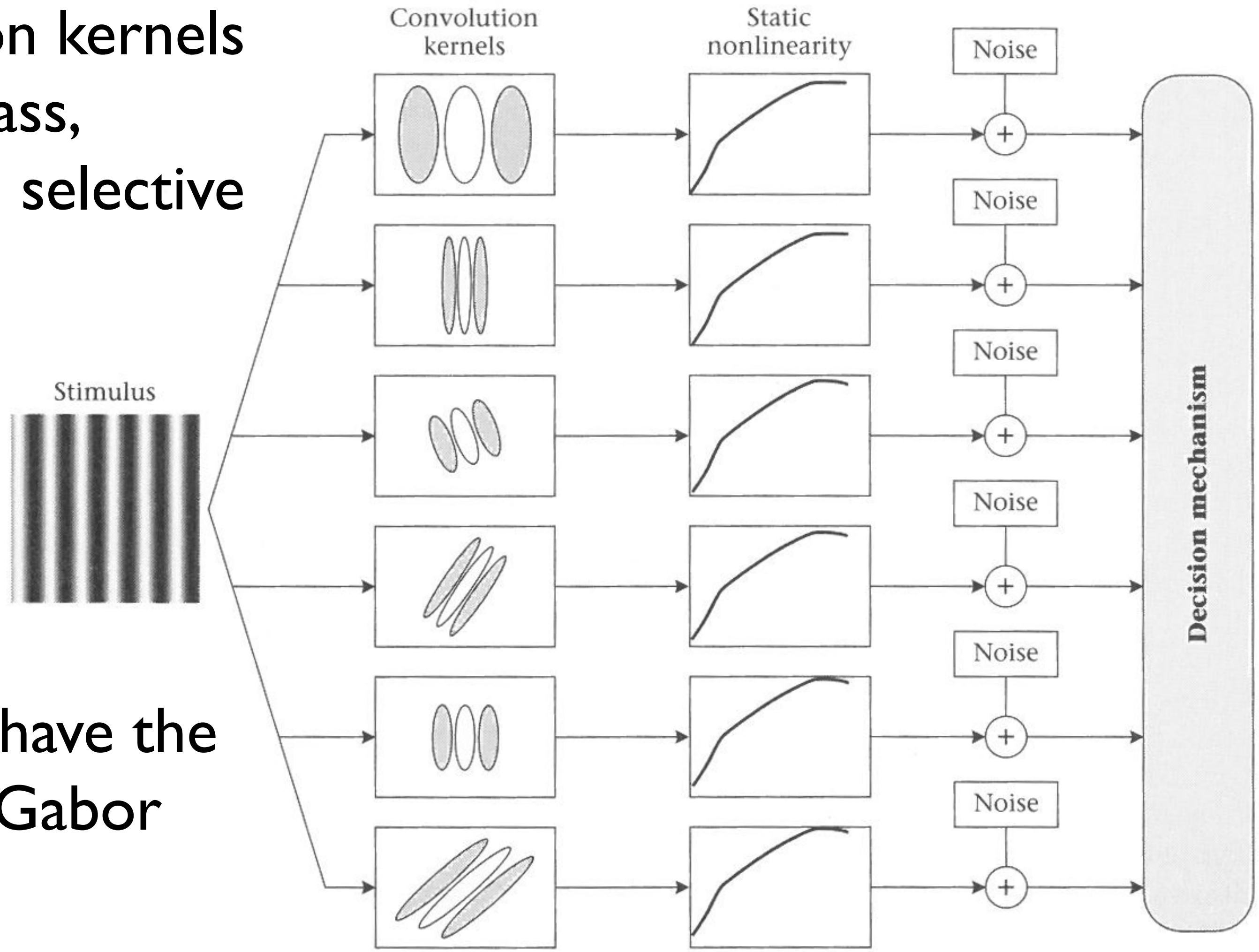
- The visual information is passed to the visual cortex in multiple channels
  - The channels are selective to spatial frequency, temporal frequency and orientation
  - Each channel is affected by different „noise" level
  - The CSF is the net result of information being passed in noise-affected visual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



From: Wandell, 1995

Multi-resolution visual model

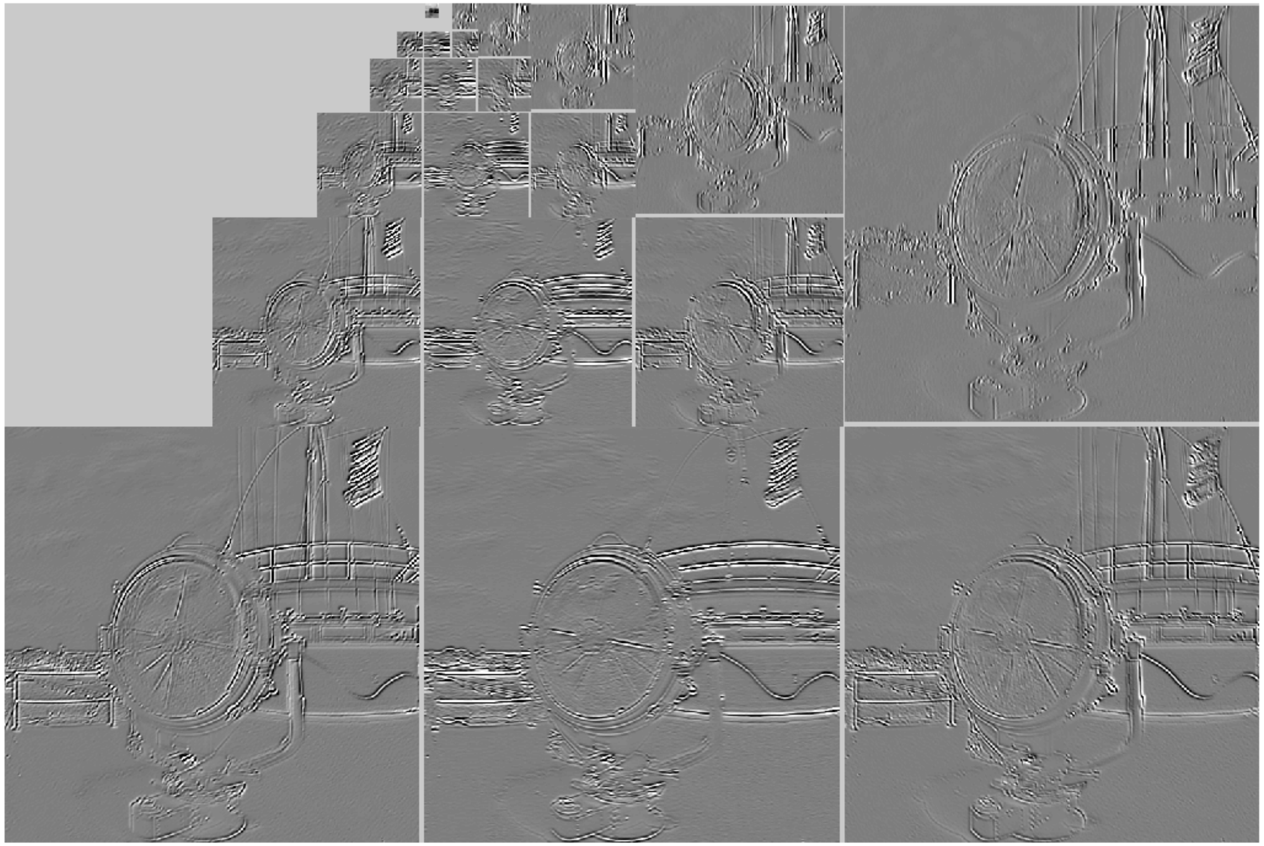
- Convolution kernels are band-pass, orientation selective filters
- The filters have the shape of a Gabor function



From: Wandell, 1995

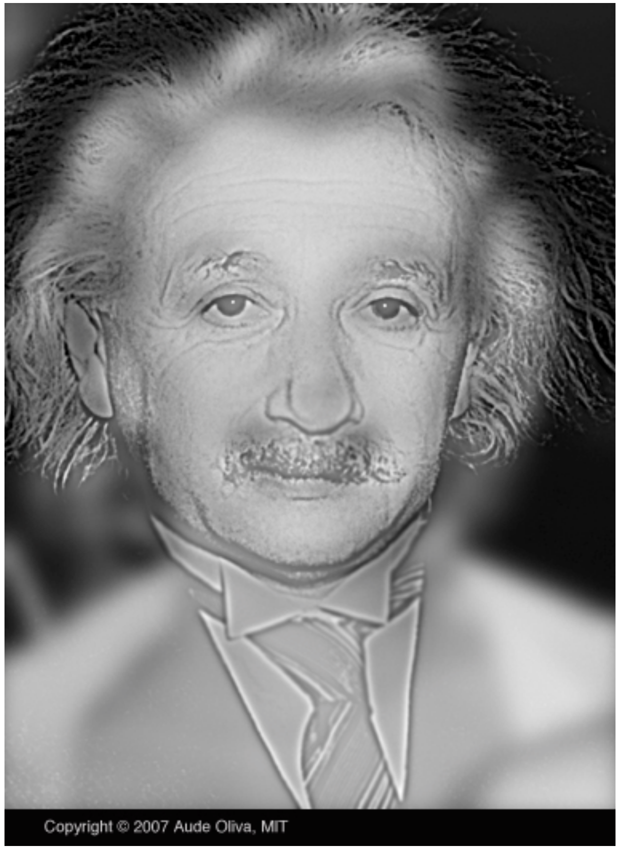
Multi-scale decomposition

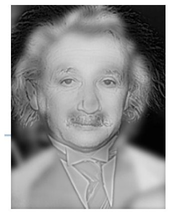
Steerable pyramid decomposition



Applications of multi-scale models

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

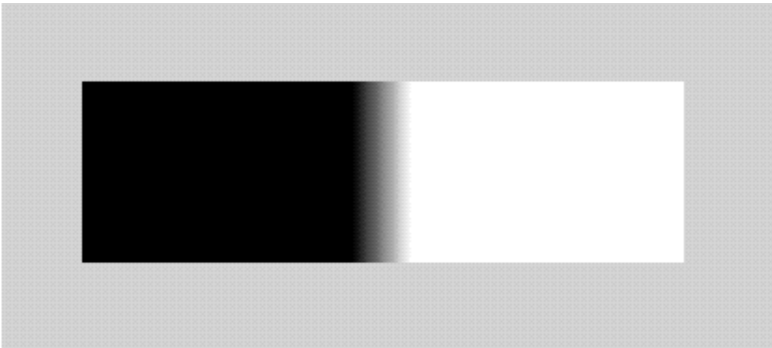


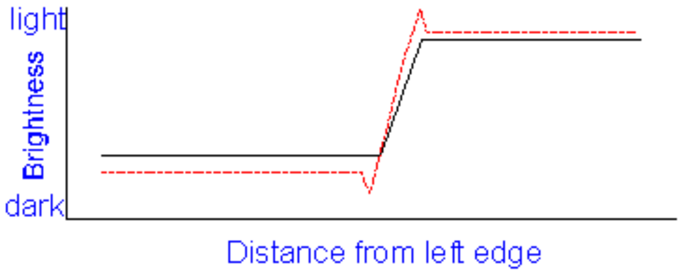


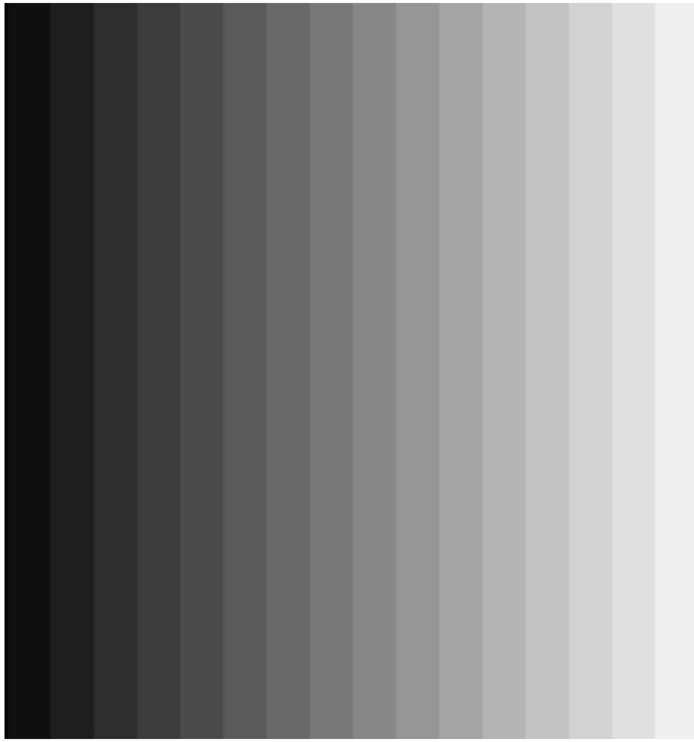
Hybrid Images by Aude Oliva
http://cvcl.mit.edu/hybrid_gallery

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“



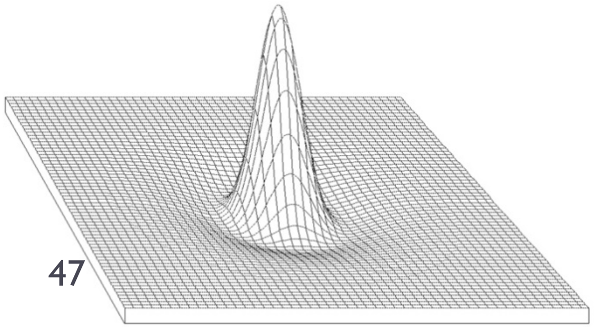


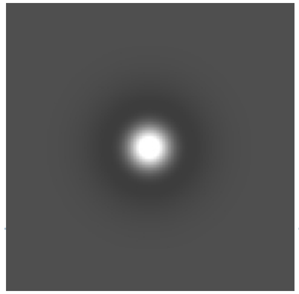


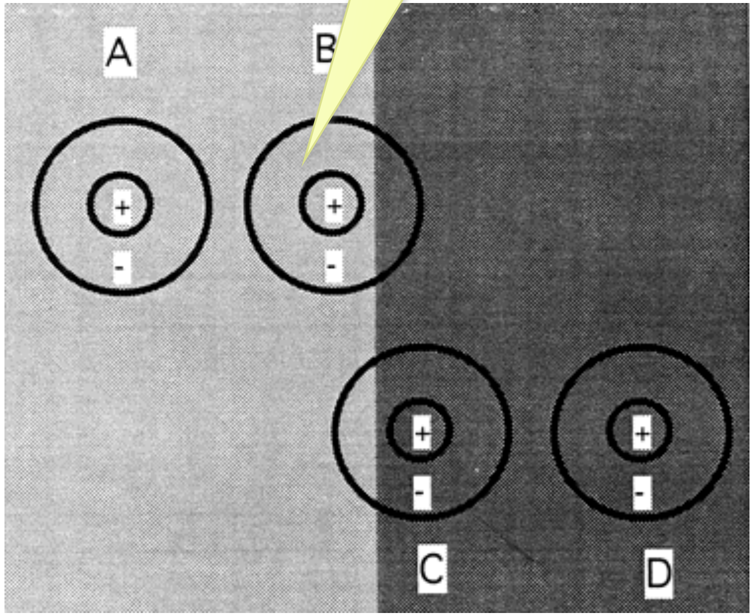
Centre-surround (Lateral Inhibition)

- "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)

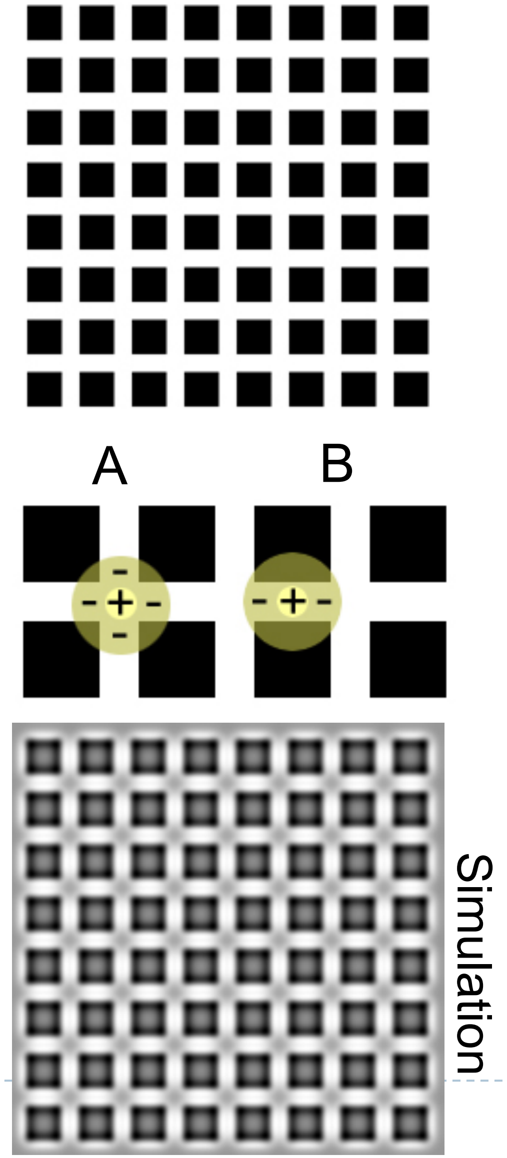


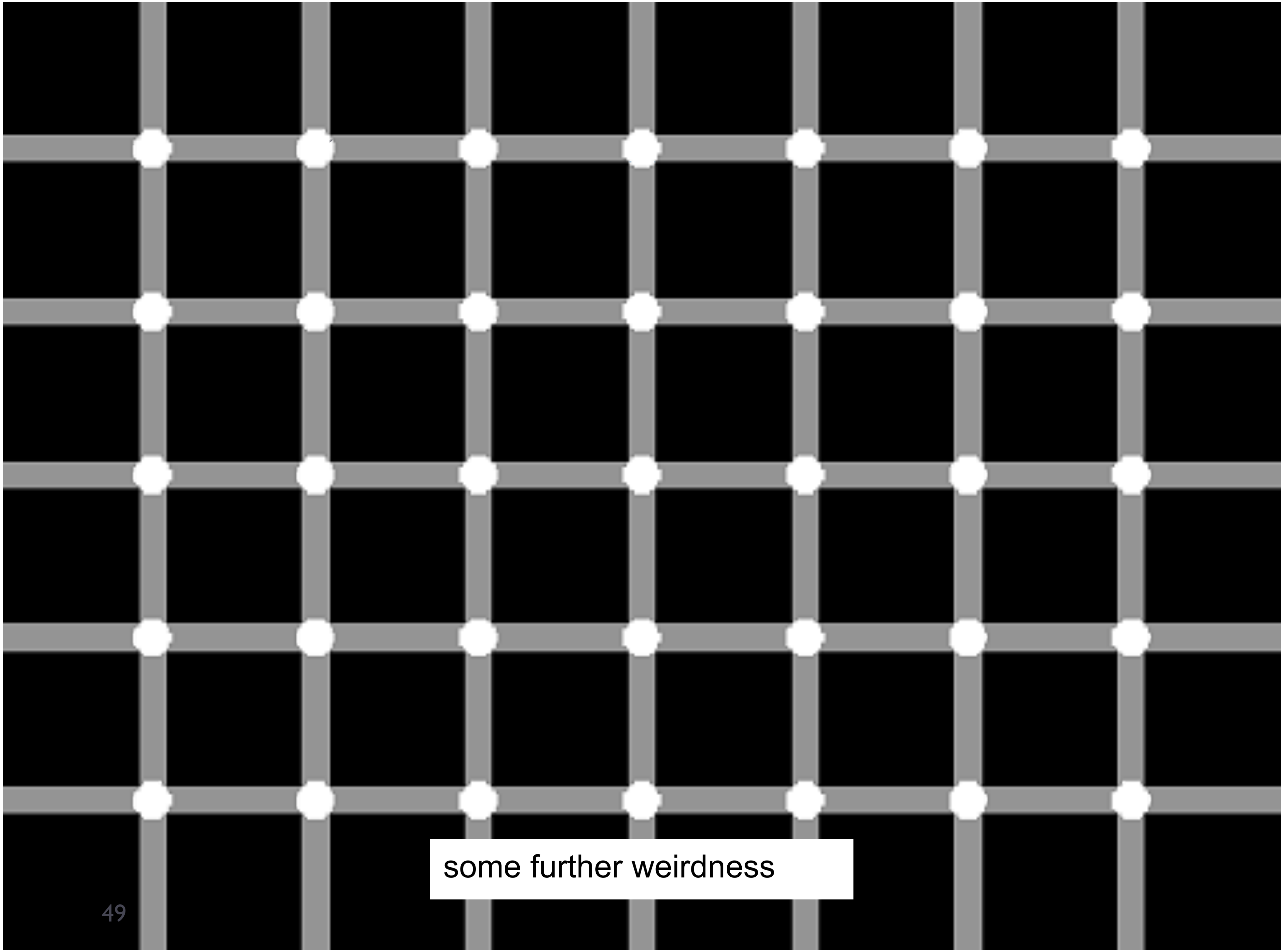




Centre-surround: Hermann Grid

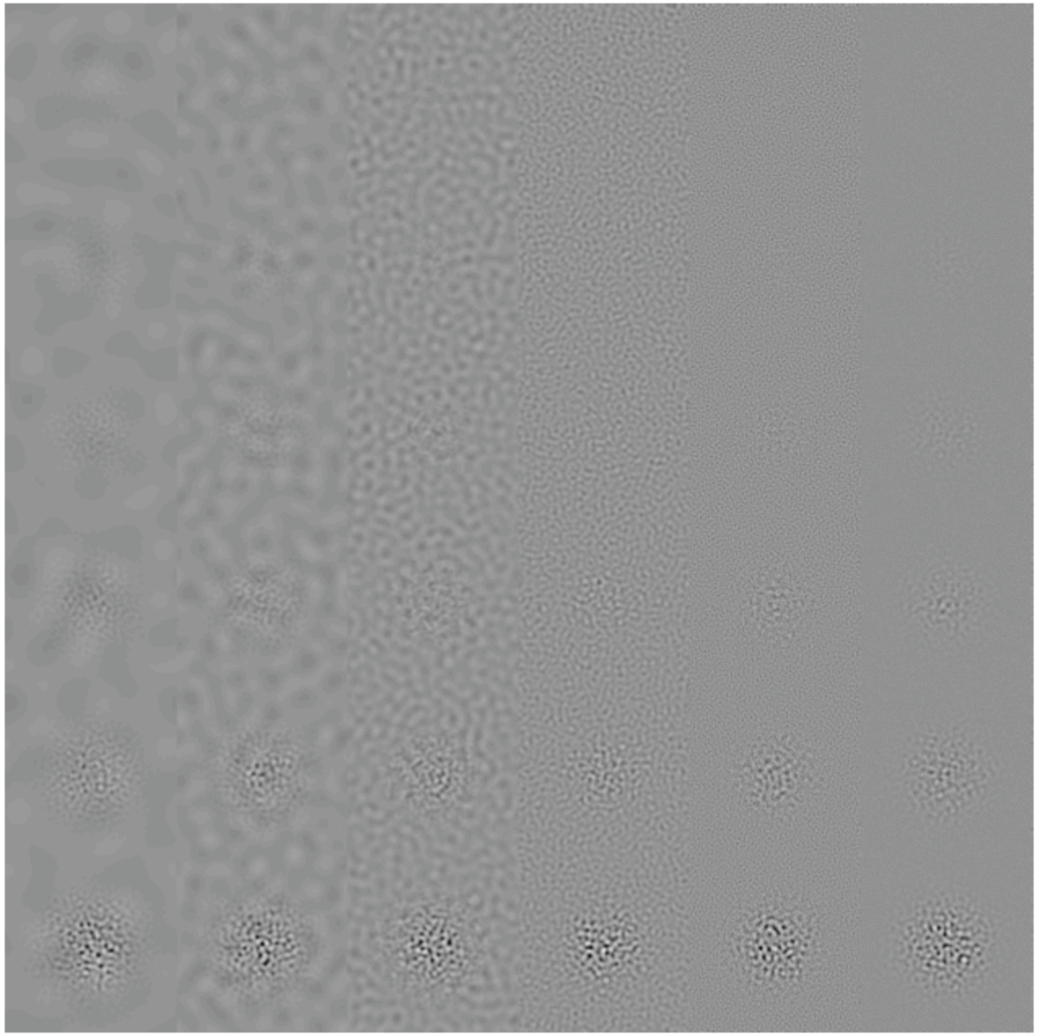
- Dark dots at crossings
- Explanation
  - Crossings (A)
    - More surround stimulation (more bright area)
    - ⇒ Less inhibition
    - ⇒ Weaker response
  - Streets (B)
    - Less surround stimulation
    - ⇒ More inhibition
    - ⇒ Greater response
- Simulation
  - Darker at crossings, brighter in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



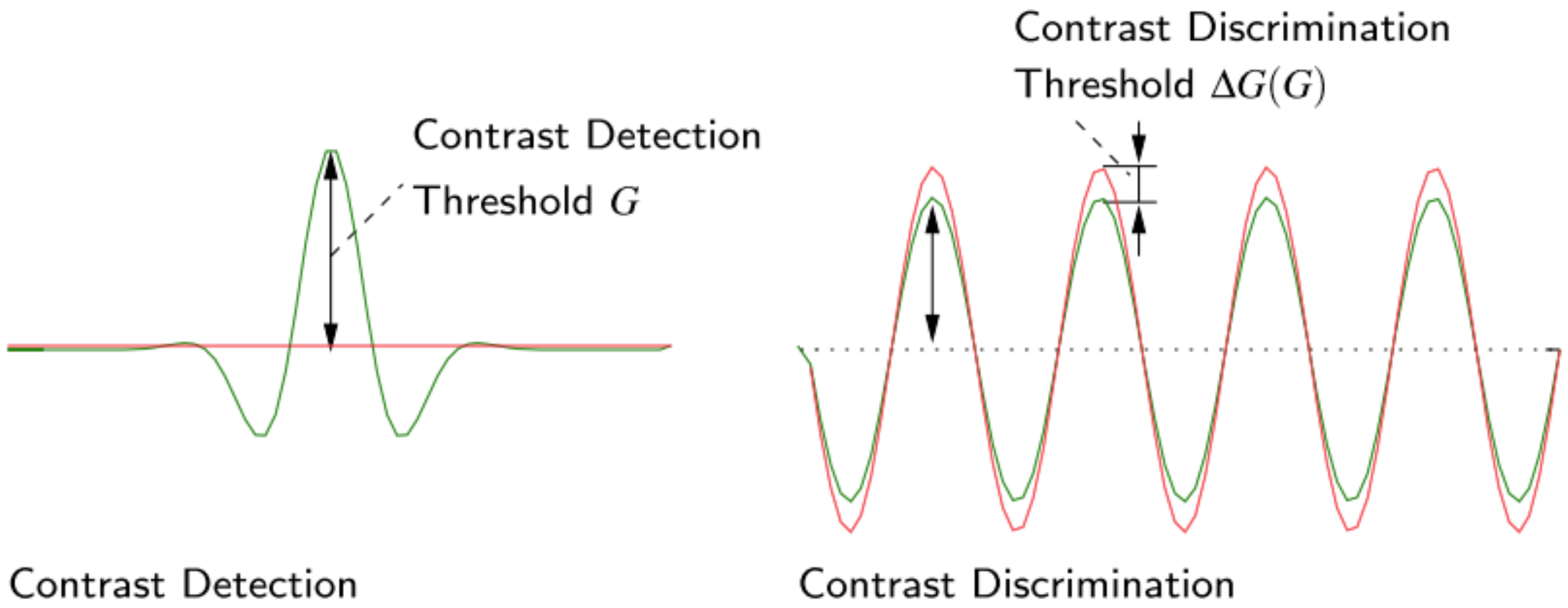

some further weirdness

Contrast masking

In which vertical bar is the pattern the most and the least visible?



Detection vs. discrimination



Definitions of contrast

Simple Contrast

$$C_s = \frac{L_{max}}{L_{min}}$$

Weber Fraction

$$W = \frac{\Delta L}{L_{min}}$$

Logarithmic Ratio

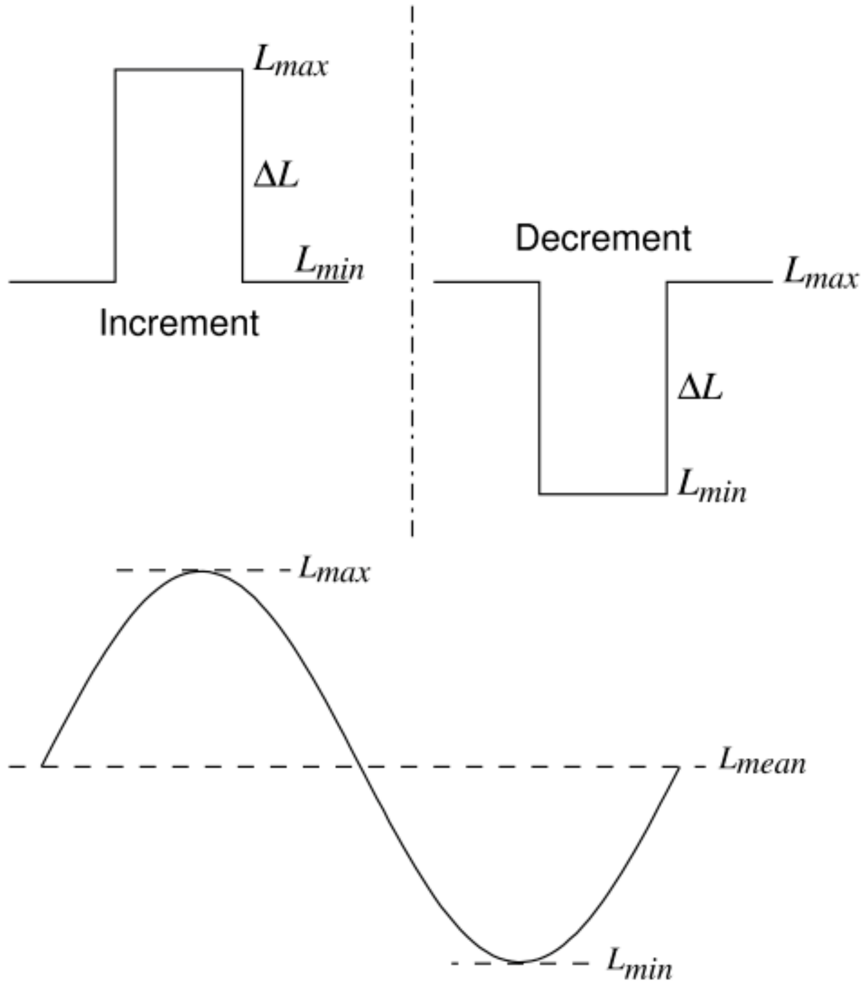
$$G = \log_{10}\left(\frac{L_{max}}{L_{min}}\right)$$

Michelson Contrast

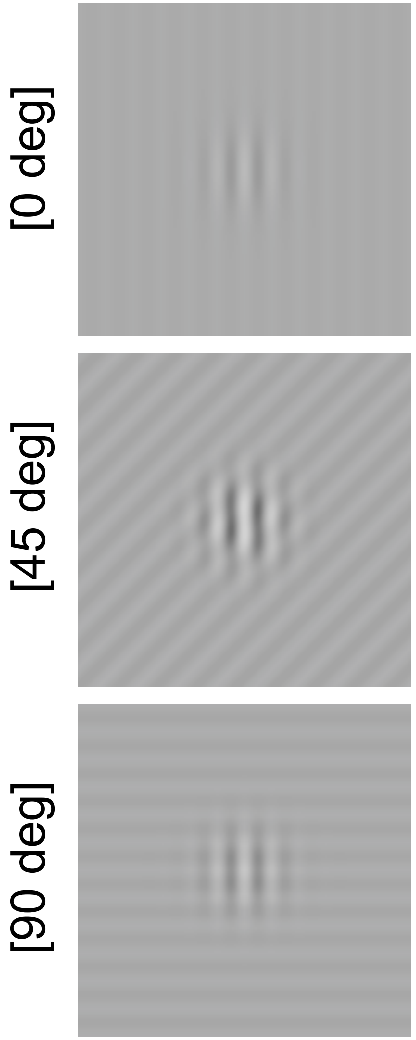
$$M = \frac{|L_{max} - L_{min}|}{L_{max} + L_{min}}$$

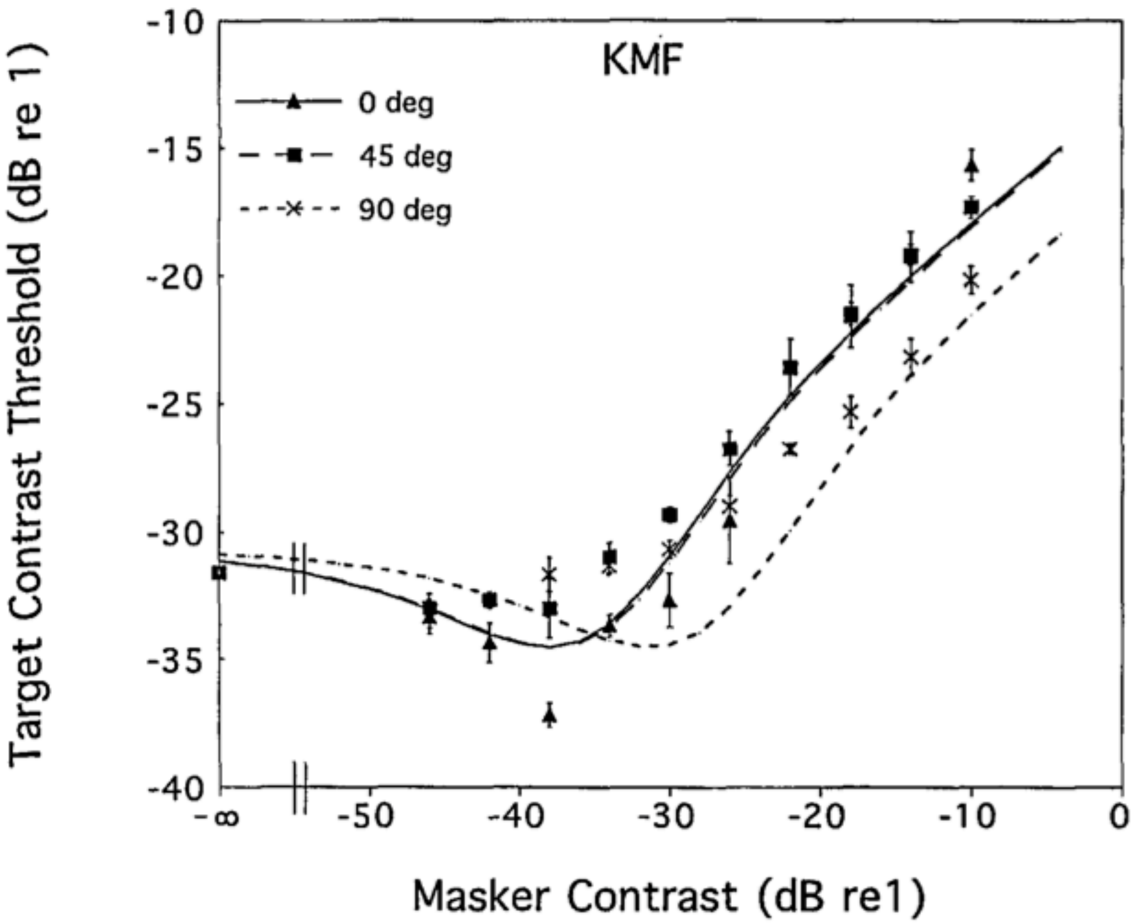
Signal to Noise Ratio

$$SNR = 20 \cdot \log_{10}\left(\frac{L_{max}}{L_{min}}\right)$$



Visual masking experiments - discrimination

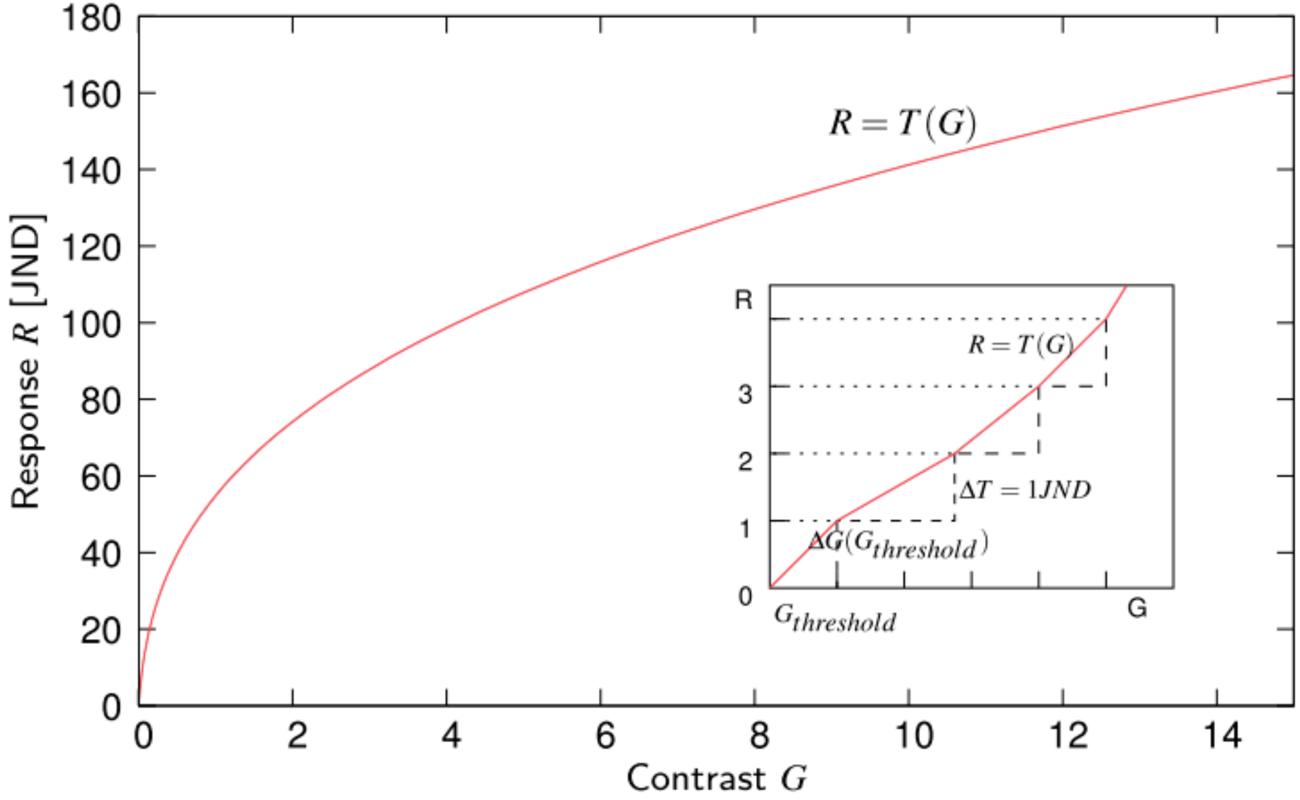




[Foley, JOSA A, 1994]

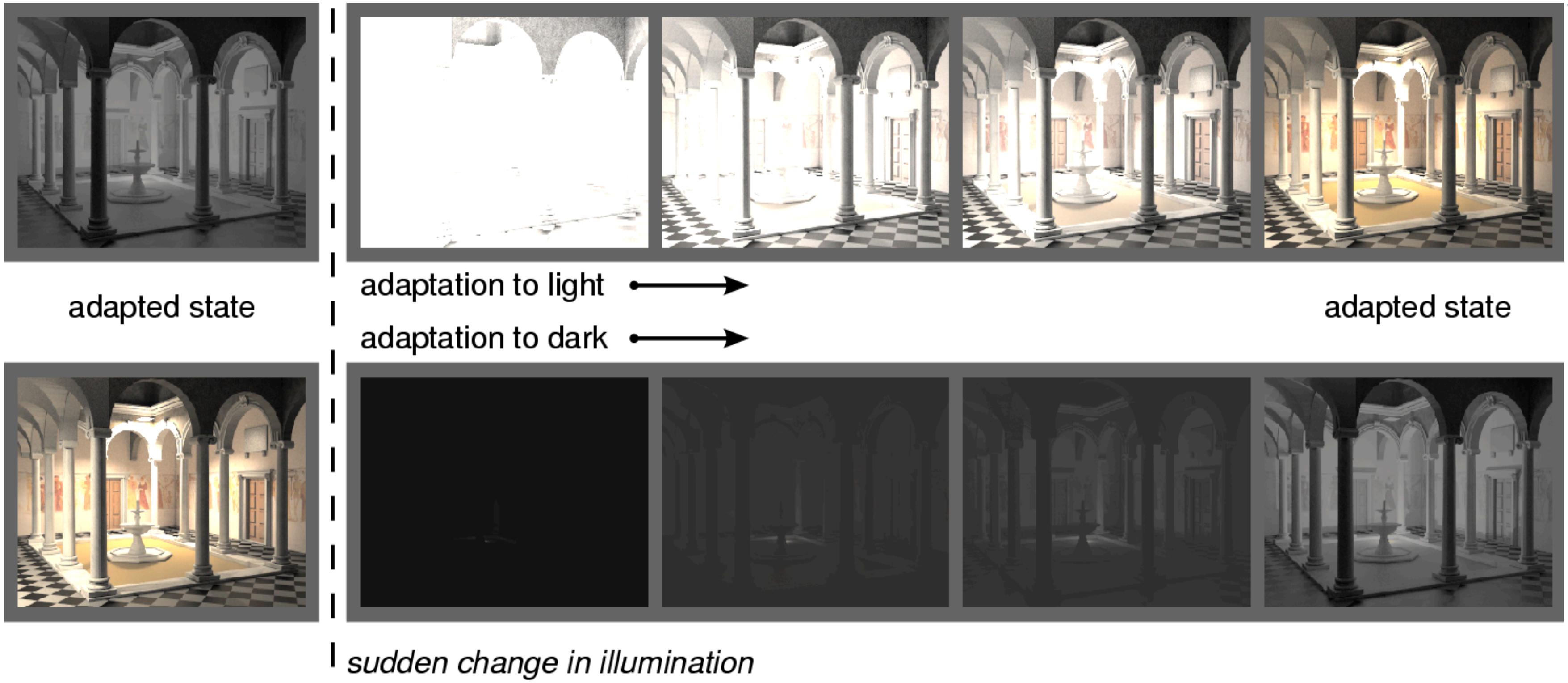
Contrast transducer

- Similar to JND encoding
- Trans. from physical contrast into "perceived" contrast



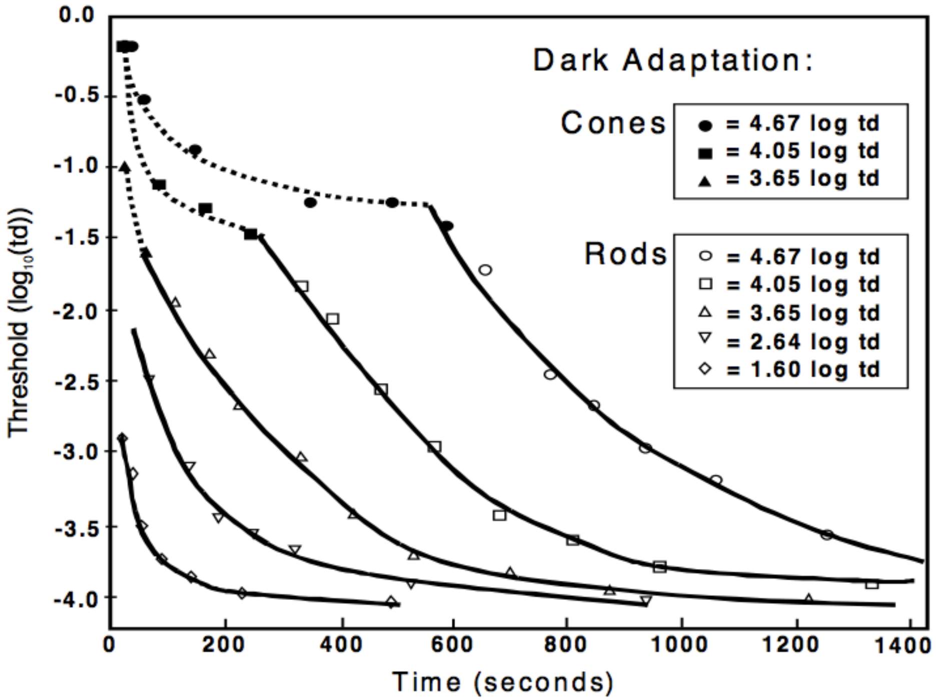
Light and dark adaptation

Light and dark adaptation

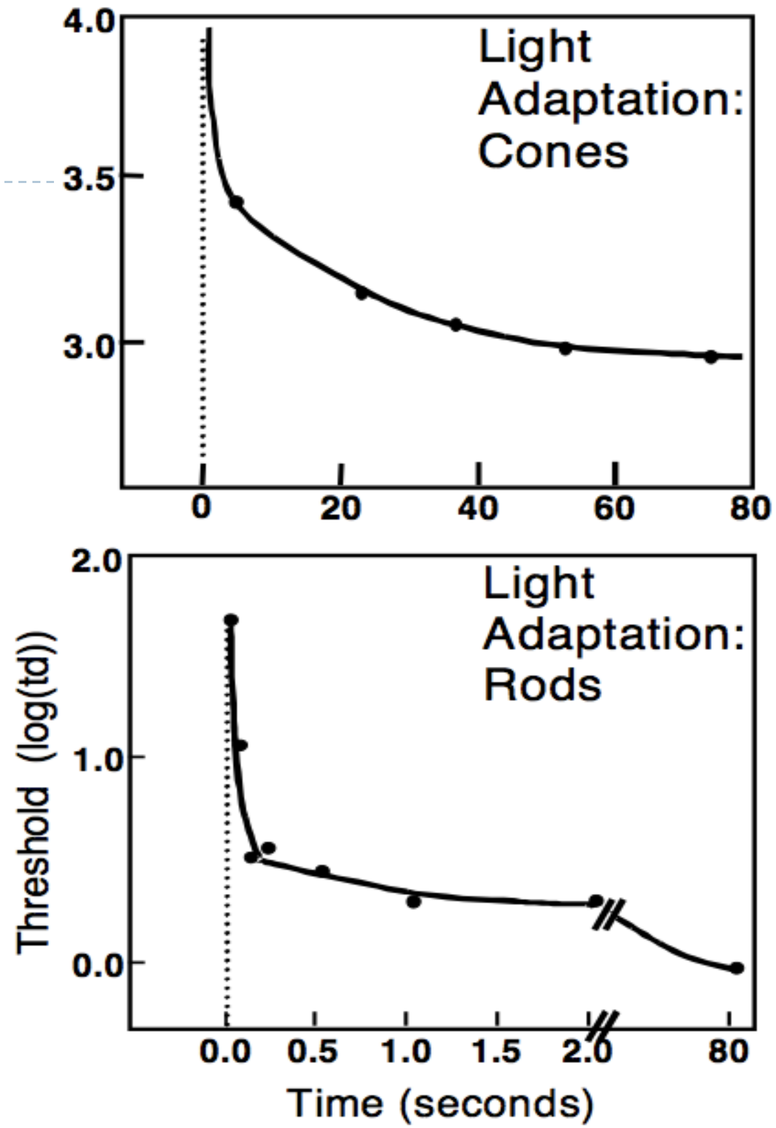


- Light adaptation: from dark to bright
- Dark adaptation: from bright to dark (much slower)

Time-course of adaptation



Bright -> Dark



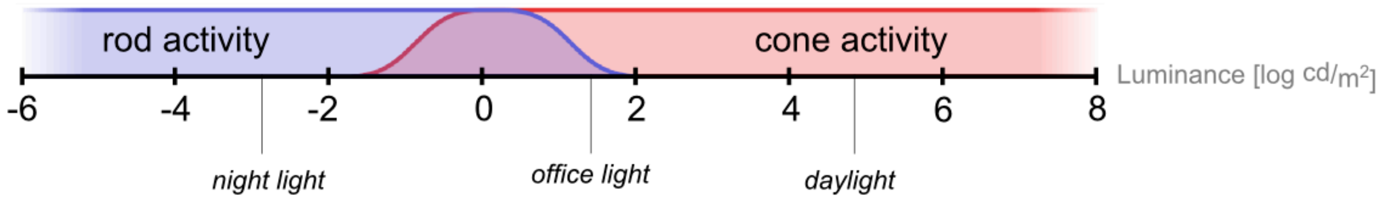
Dark -> Bright

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)

Night and daylight vision

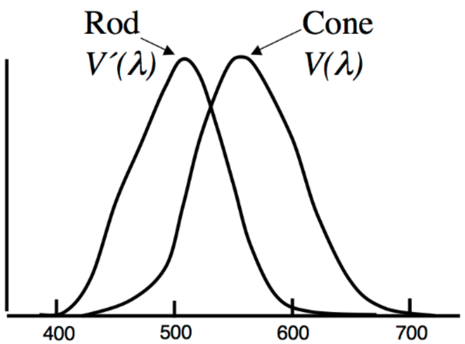
Vision mode: SCOTOPIC MESOPIC PHOTOPIC



Mode properties: monochromatic vision
limited visual acuity

good color perception
good visual acuity

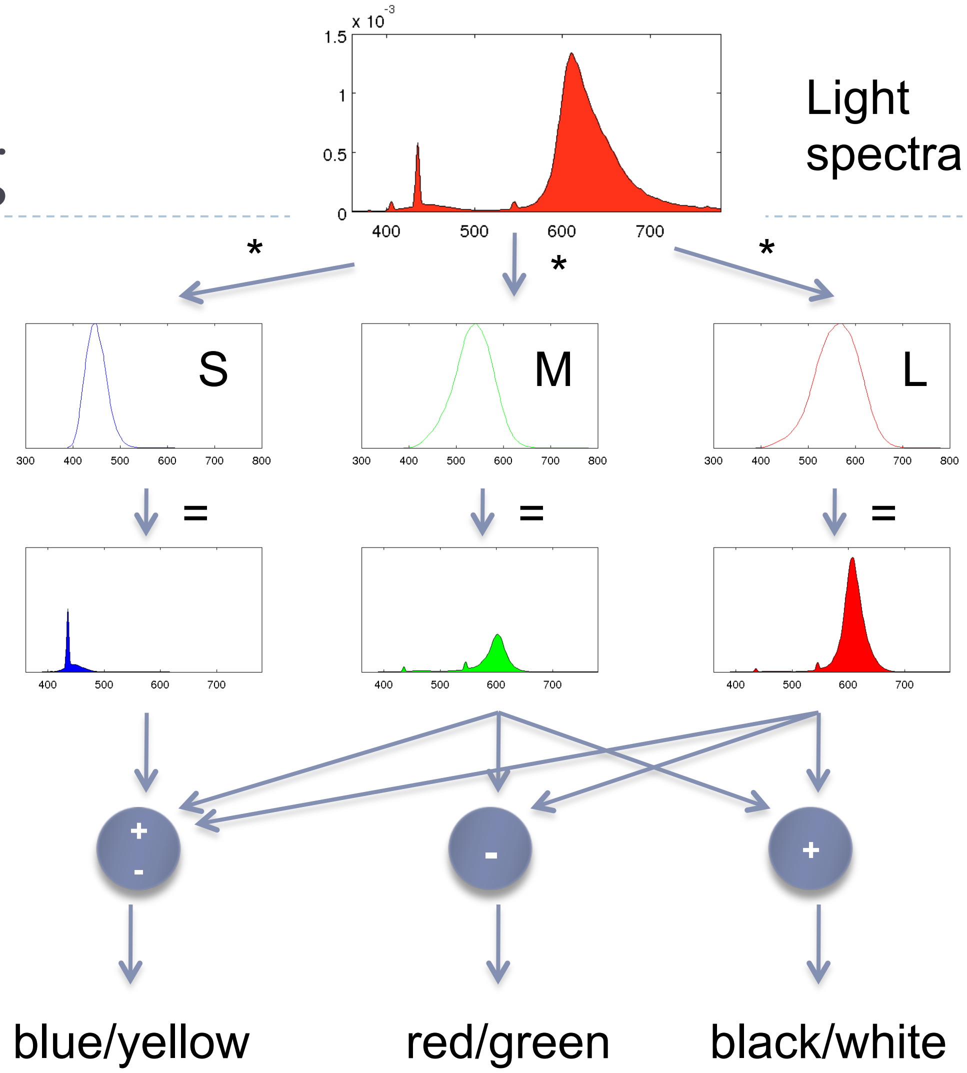
Luminous efficiency



Opponent colours and spatial colour vision

Colour processing

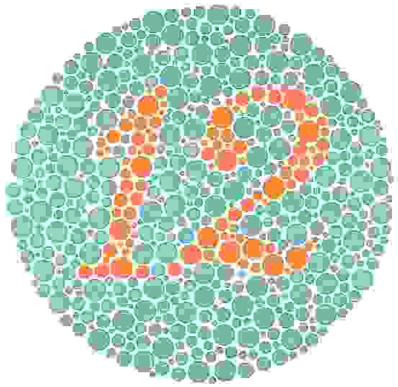
- Light is sensed by L, M and S, cones
  - Each cone type is sensitive to different wavelengths
- Responses from L, M and S cones are combined into three opponent pathways
  - achromatic (black/ white) pathway – luminance
  - 2 colour opponent pathways
- Rationale: improve coding efficiency for natural scenes

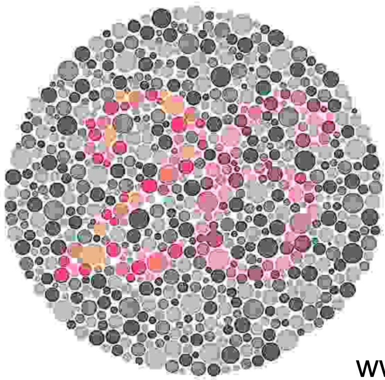


Colour perception

- Di-chromaticity (dogs, cats)
  - Yellow & blue-violet
  - Green, orange, red indistinguishable
- Tri-chromaticity (humans, monkeys)
  - Red, green, blue
  - Colour-blindness
    - Most often men, green-red colour-blindness

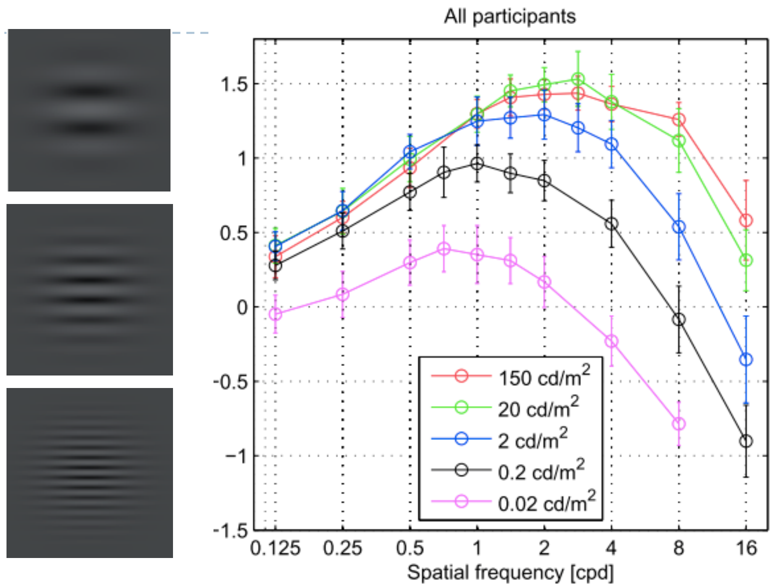
www.lam.mus.ca.us/cats/color/

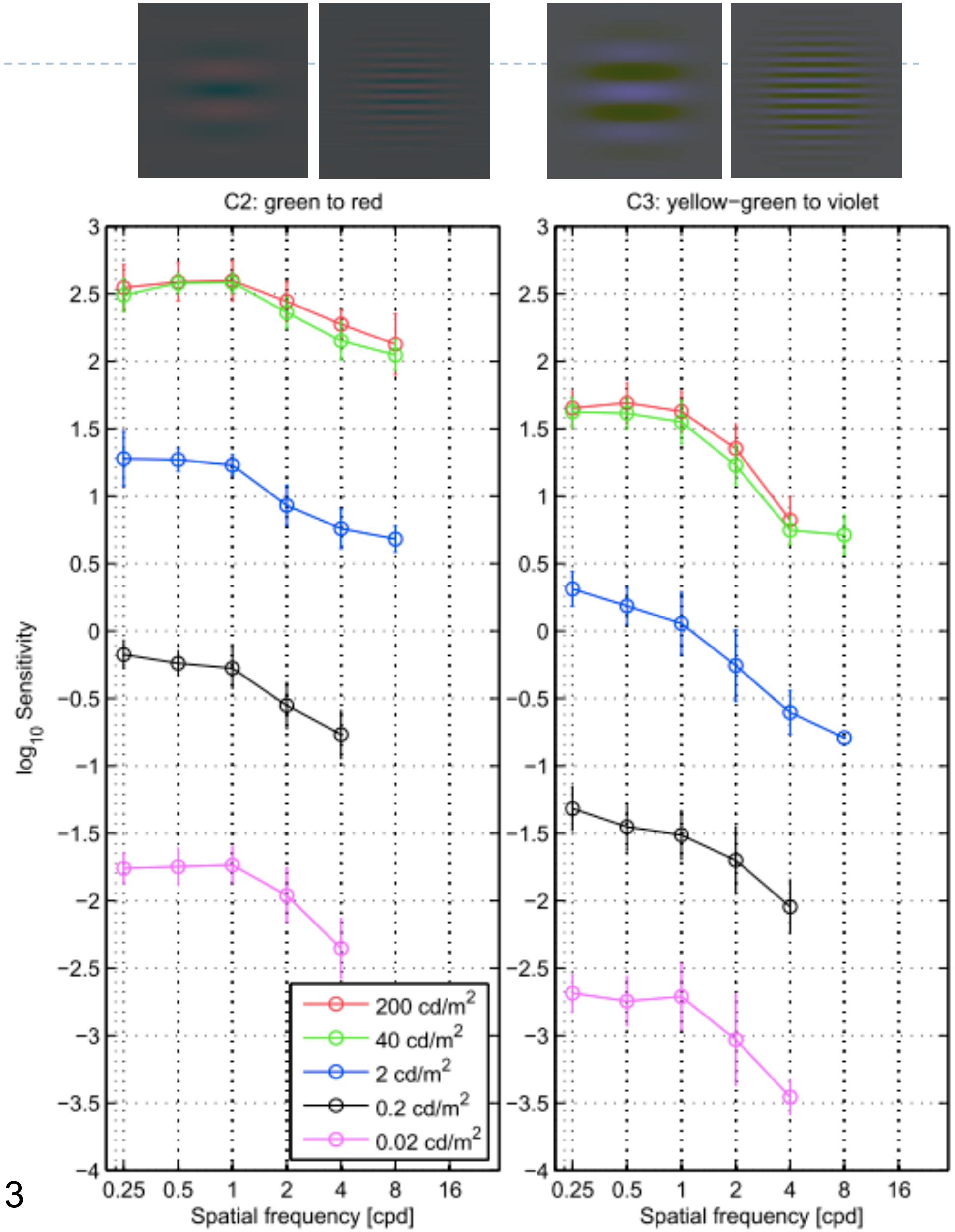




www.colorcube.com/illusions/clrblnd.html

Colour Contrast Sensitivity

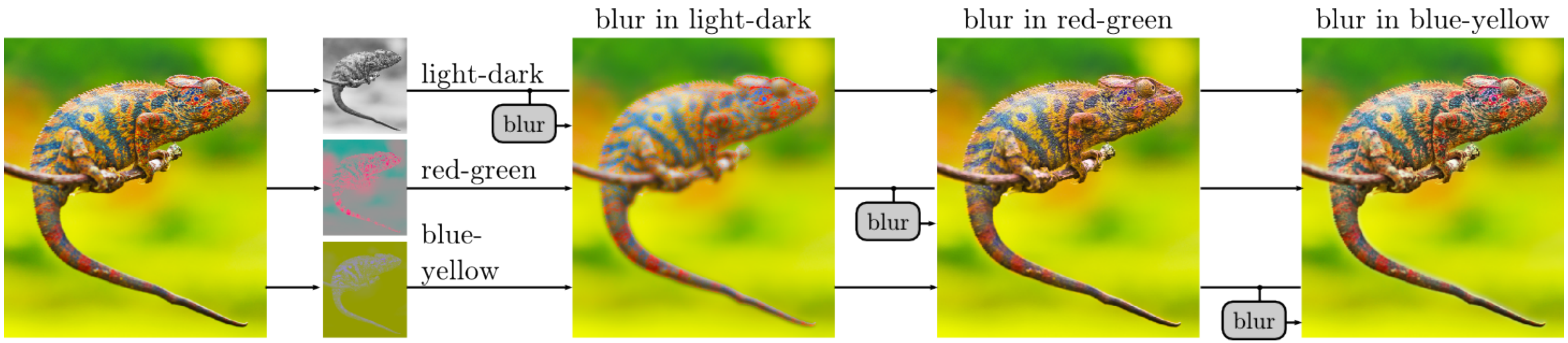




- Colour vs. luminance vision system
  - Higher sensitivity at lower frequencies for colour
  - High frequencies less visible

From: Kim et al., HVEI 2013

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)

Depth perception

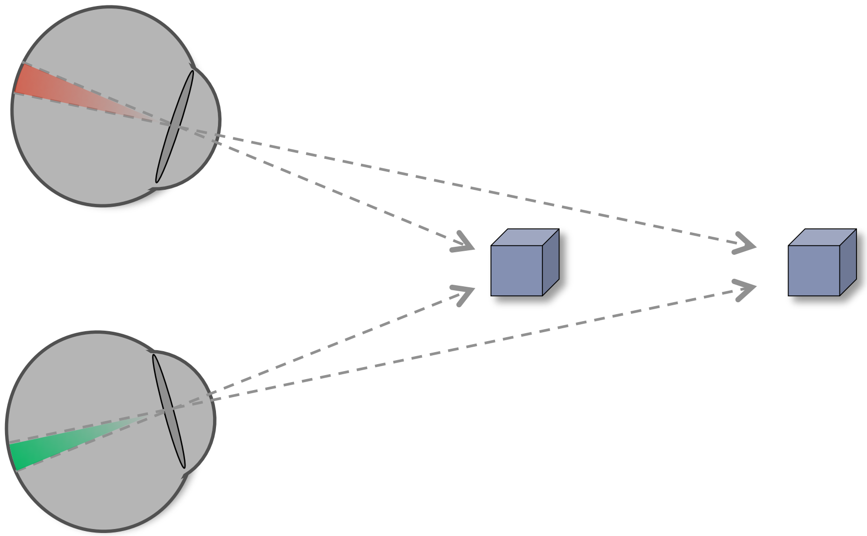
The slides in this section are the courtesy of Piotr Didyk (http://people.mpi-inf.mpg.de/~pdidyk/)

Depth perception

We see depth due to depth cues.

Stereoscopic depth cues:

binocular disparity



Depth perception

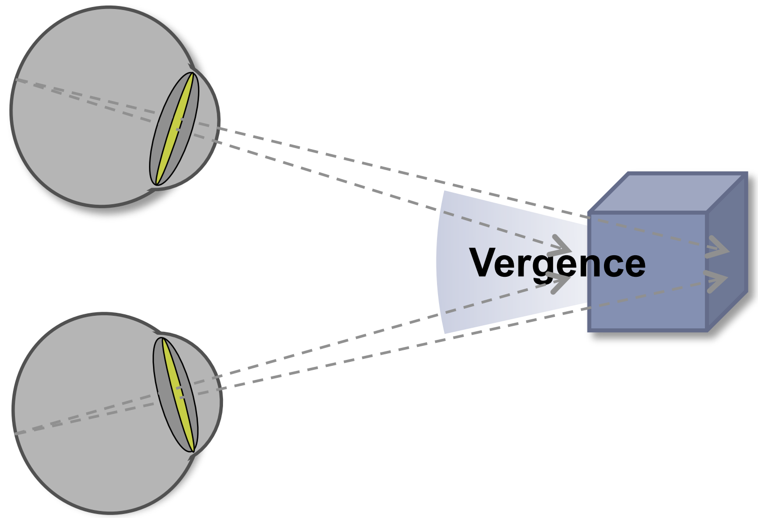
We see depth due to depth cues.

Stereoscopic depth cues:

binocular disparity

Ocular depth cues:

accommodation, vergence



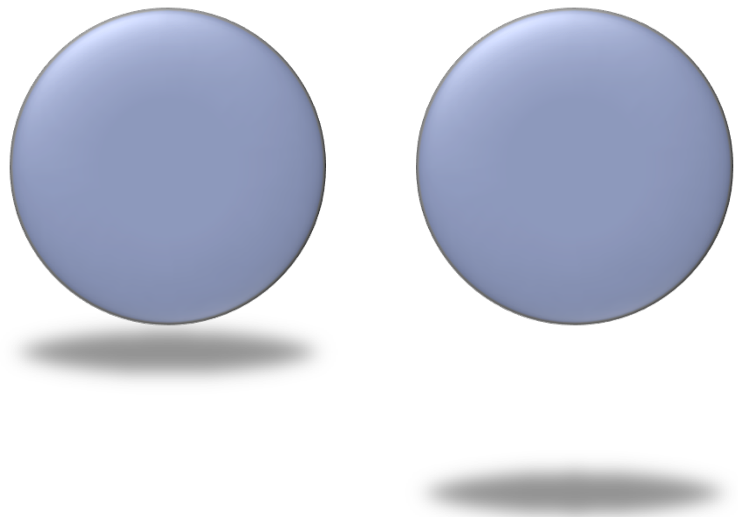
Depth perception

We see depth due to depth cues.

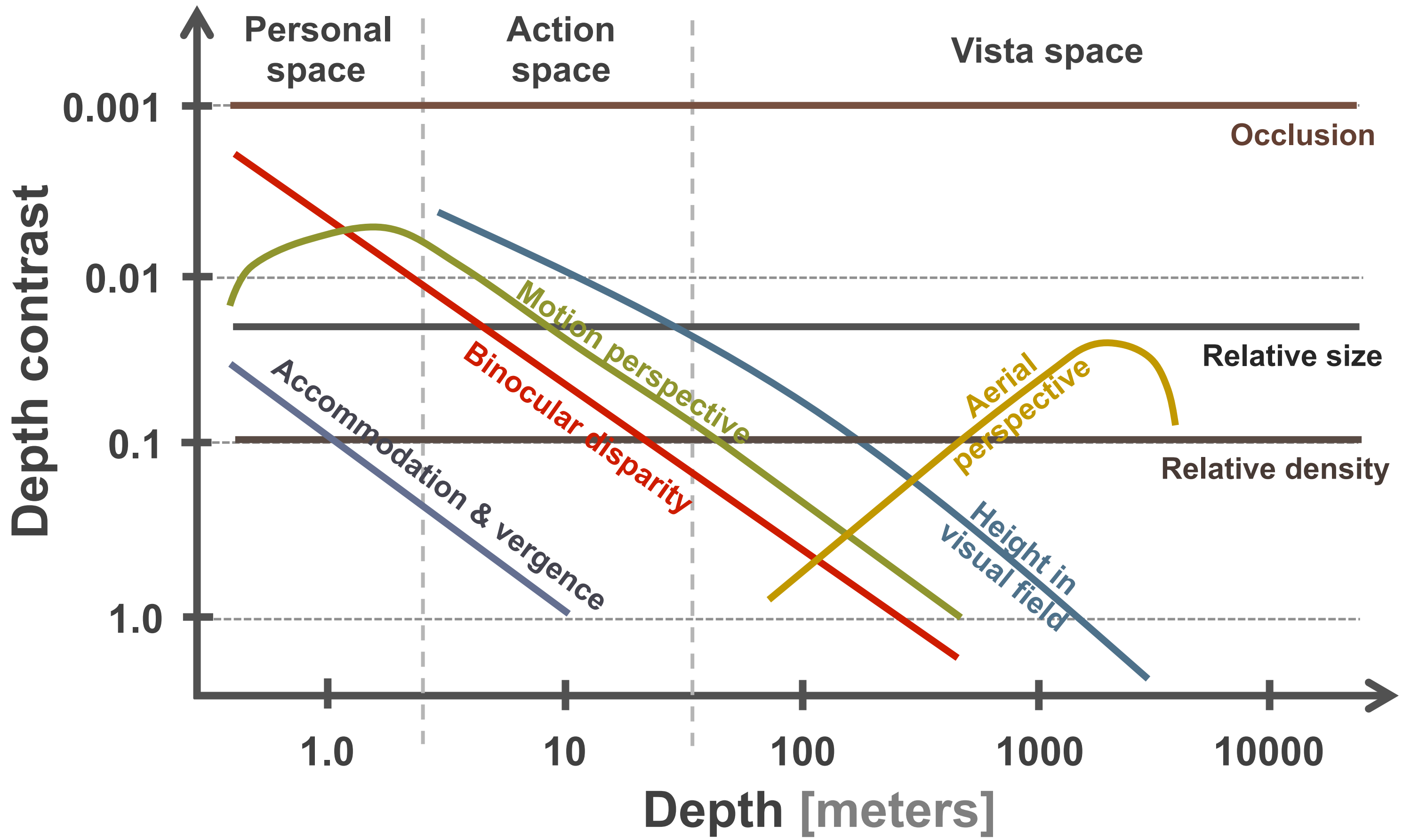
Stereoscopic depth cues:
binocular disparity

Ocular depth cues:
accommodation, vergence

Pictorial depth cues:
occlusion, size, shadows…



Cues sensitivity



"Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth"
by Cutting and Vishton [1995]

Depth perception

We see depth due to depth cues.

Stereoscopic depth cues:
binocular disparity

Ocular depth cues:
accommodation, vergence

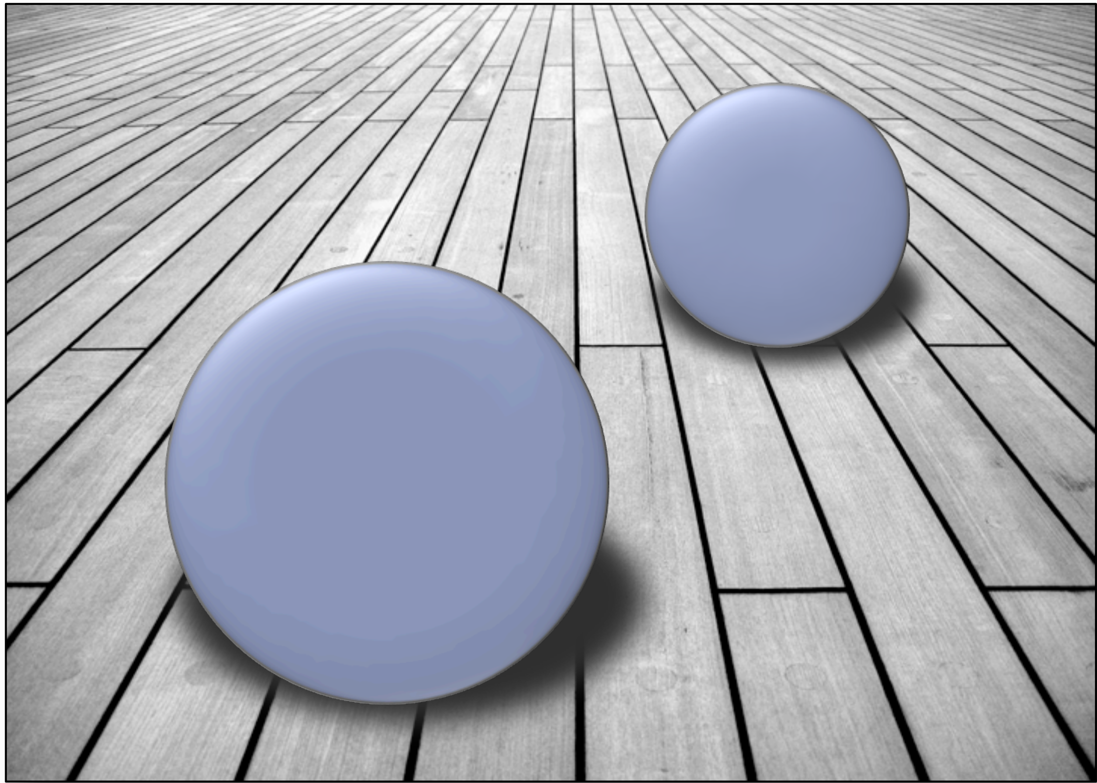
Pictorial depth cues:
occlusion, size, shadows…

Challenge:
Consistency is required!

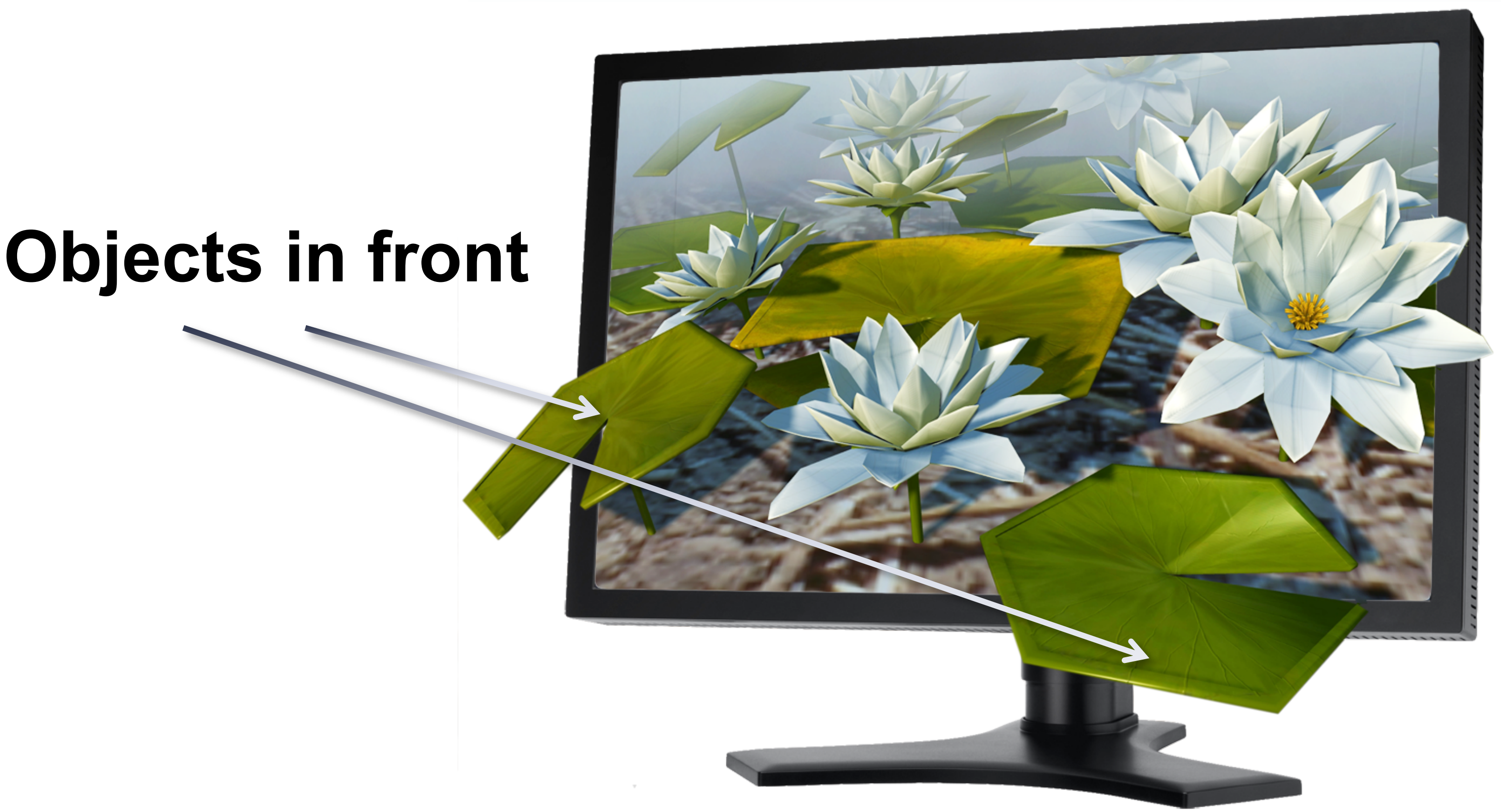
Simple conflict example

Present cues:

- Size
- Shadows
- Perspective
- **Occlusion**



Disparity & occlusion conflict



Disparity & occlusion conflict

Depth perception

We see depth due to depth cues.

Stereoscopic depth cues:
binocular disparity

Ocular depth cues:
accommodation, vergence

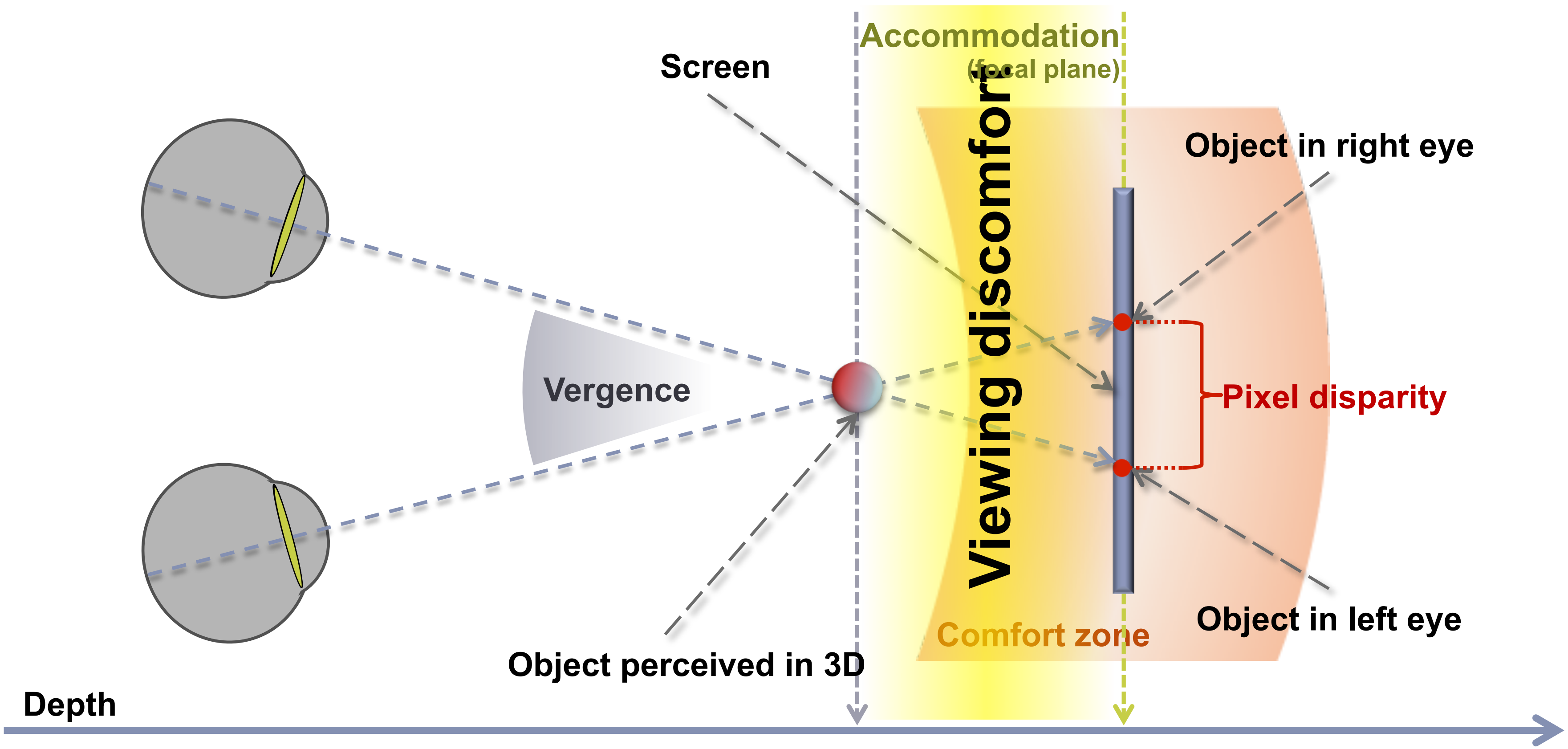
→ **Require 3D space**

We cheat our Human Visual System!

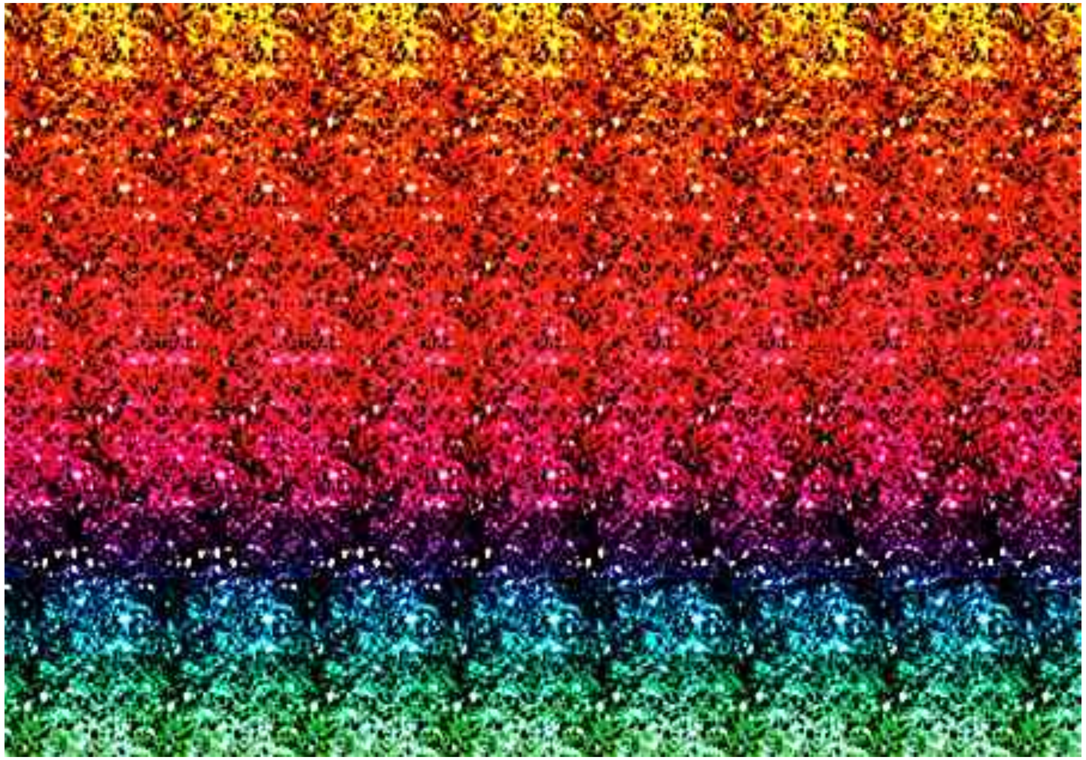
Pictorial depth cues:
occlusion, size, shadows…

→ **Reproducible on a flat displays**

Cheating our HVS

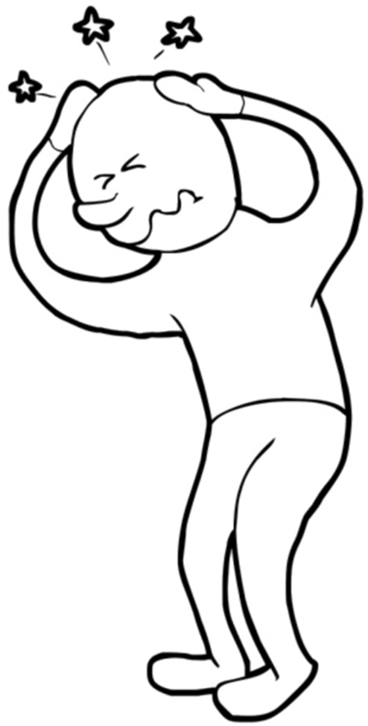


Single Image Random Dot Stereograms



- Fight the vergence vs. accommodation conflict to see the hidden image

Viewing discomfort

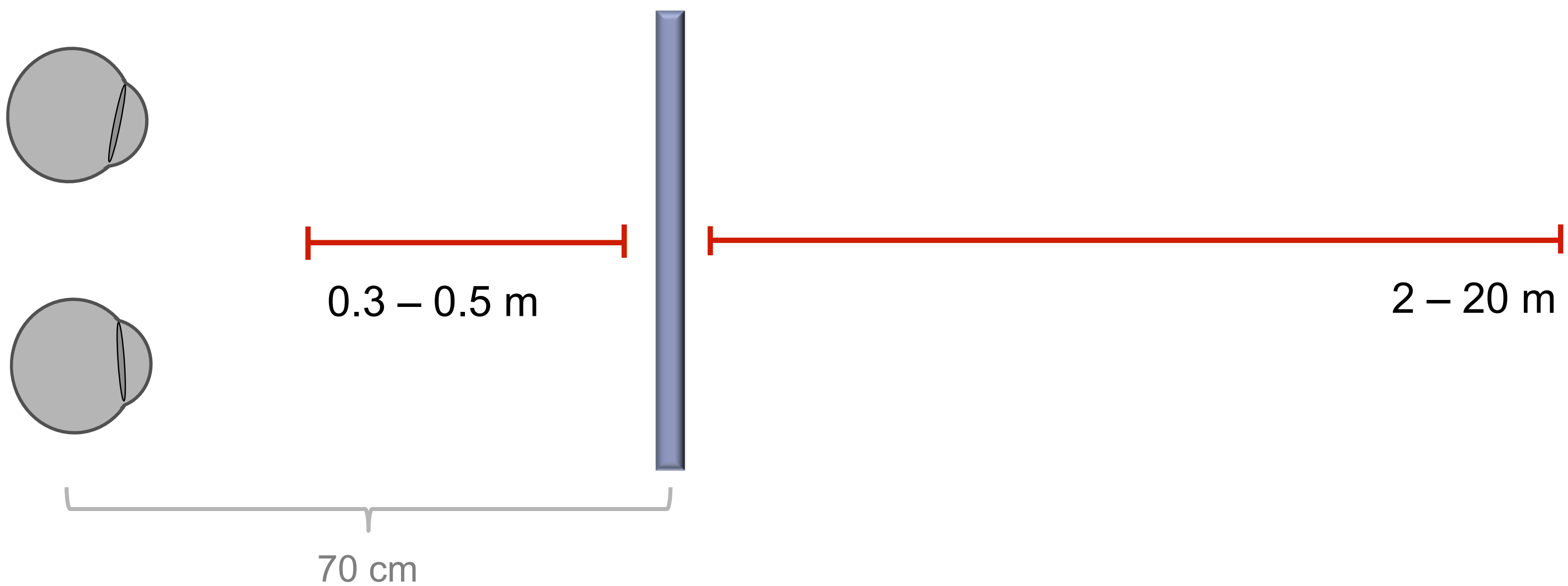


Comfort zones

Comfort zone size depends on:

- Presented content
- Viewing condition

Simple scene



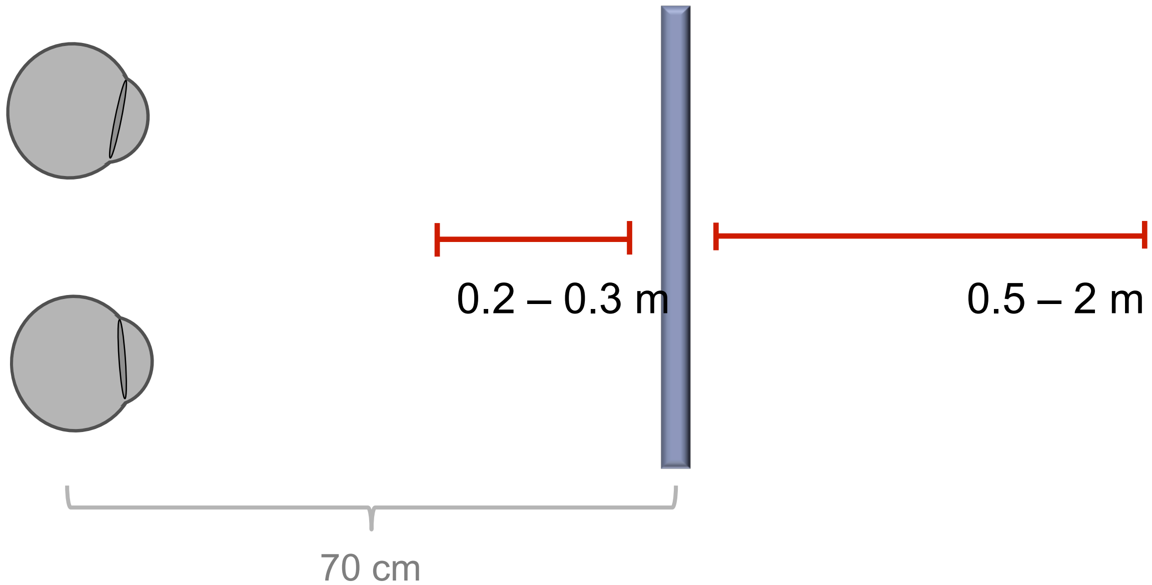
"Controlling Perceived Depth in Stereoscopic Images" by Jones et al. 2001

Comfort zones

Comfort zone size depends on:

- Presented content
- Viewing condition

Simple scene, user allowed to look away from screen



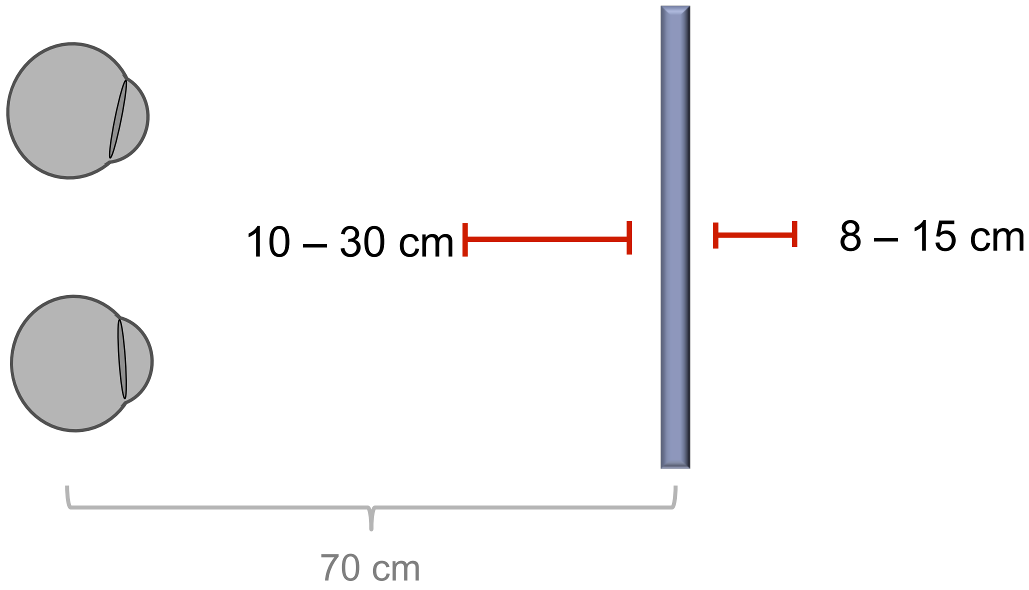
"Controlling Perceived Depth in Stereoscopic Images" by Jones et al. 2001

Comfort zones

Comfort zone size depends on:

- Presented content
- Viewing condition

Difficult scene



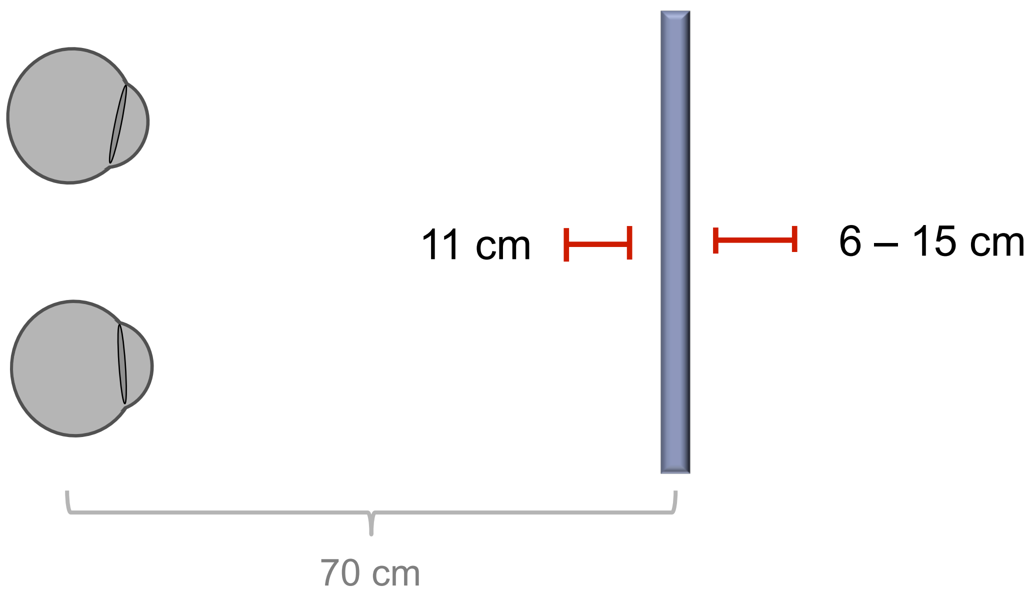
"Controlling Perceived Depth in Stereoscopic Images" by Jones et al. 2001

Comfort zones

Comfort zone size depends on:

- Presented content
- Viewing condition

Difficult scene, user allowed to look away from screen



"Controlling Perceived Depth in Stereoscopic Images" by Jones et al. 2001

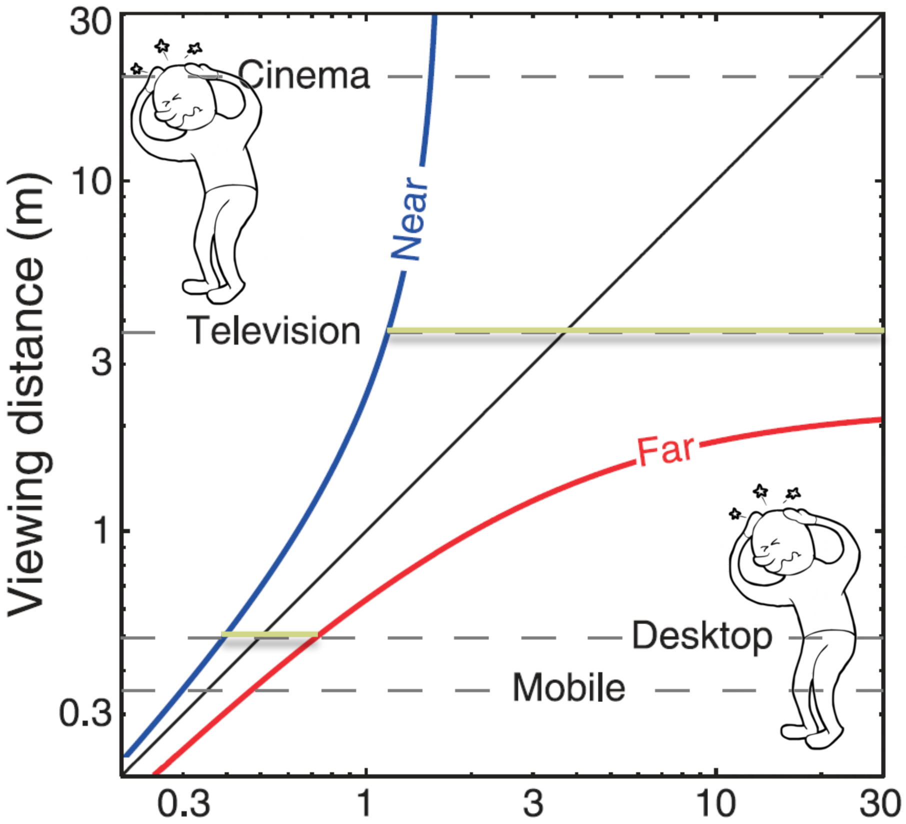
Comfort zones

Comfort zone size depends on:

- Presented content
- Viewing condition
- Screen distance

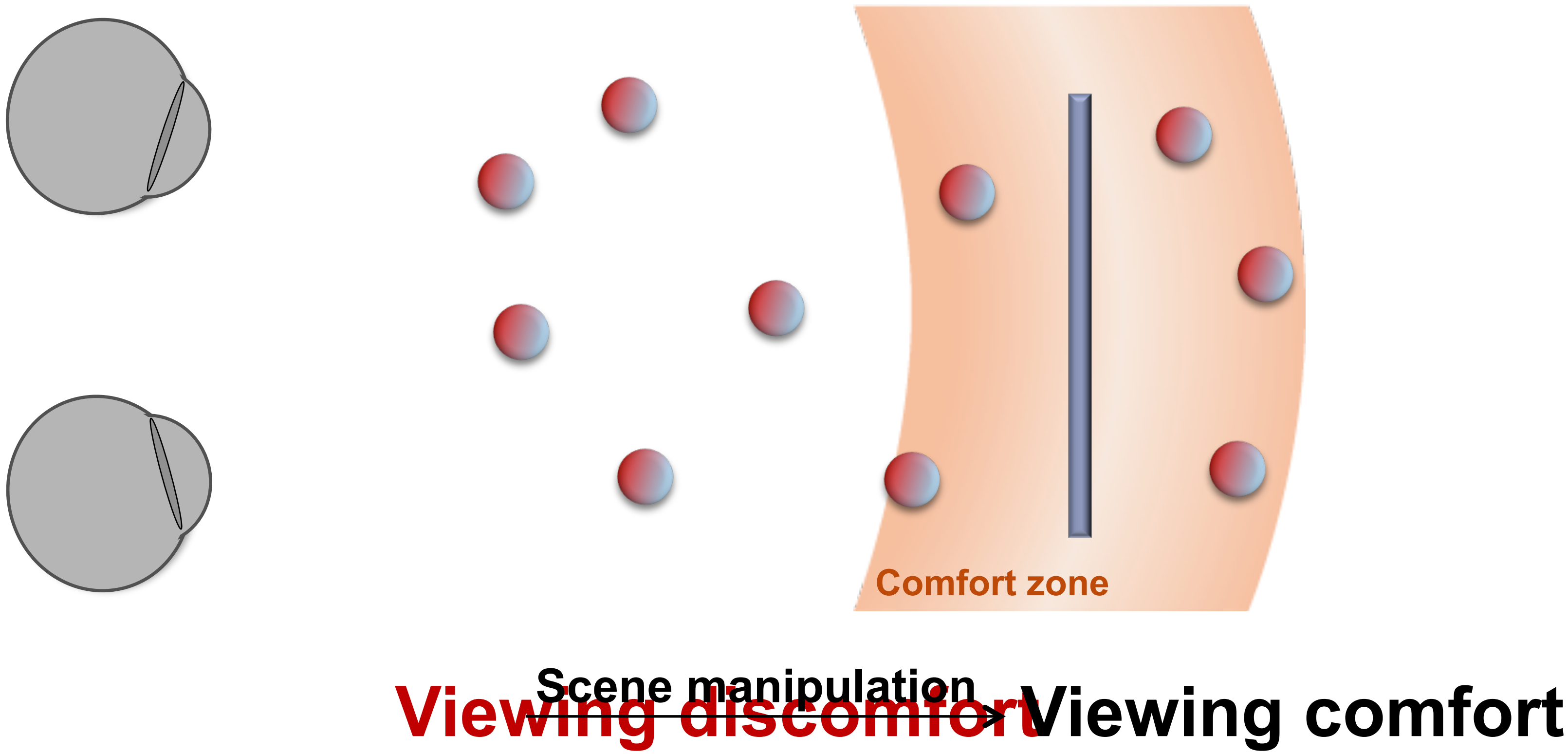
Other factors:

- Distance between eyes
- Depth of field
- Temporal coherence



"The zone of comfort: Predicting visual discomfort with stereo displays" by Shibata et al. 201

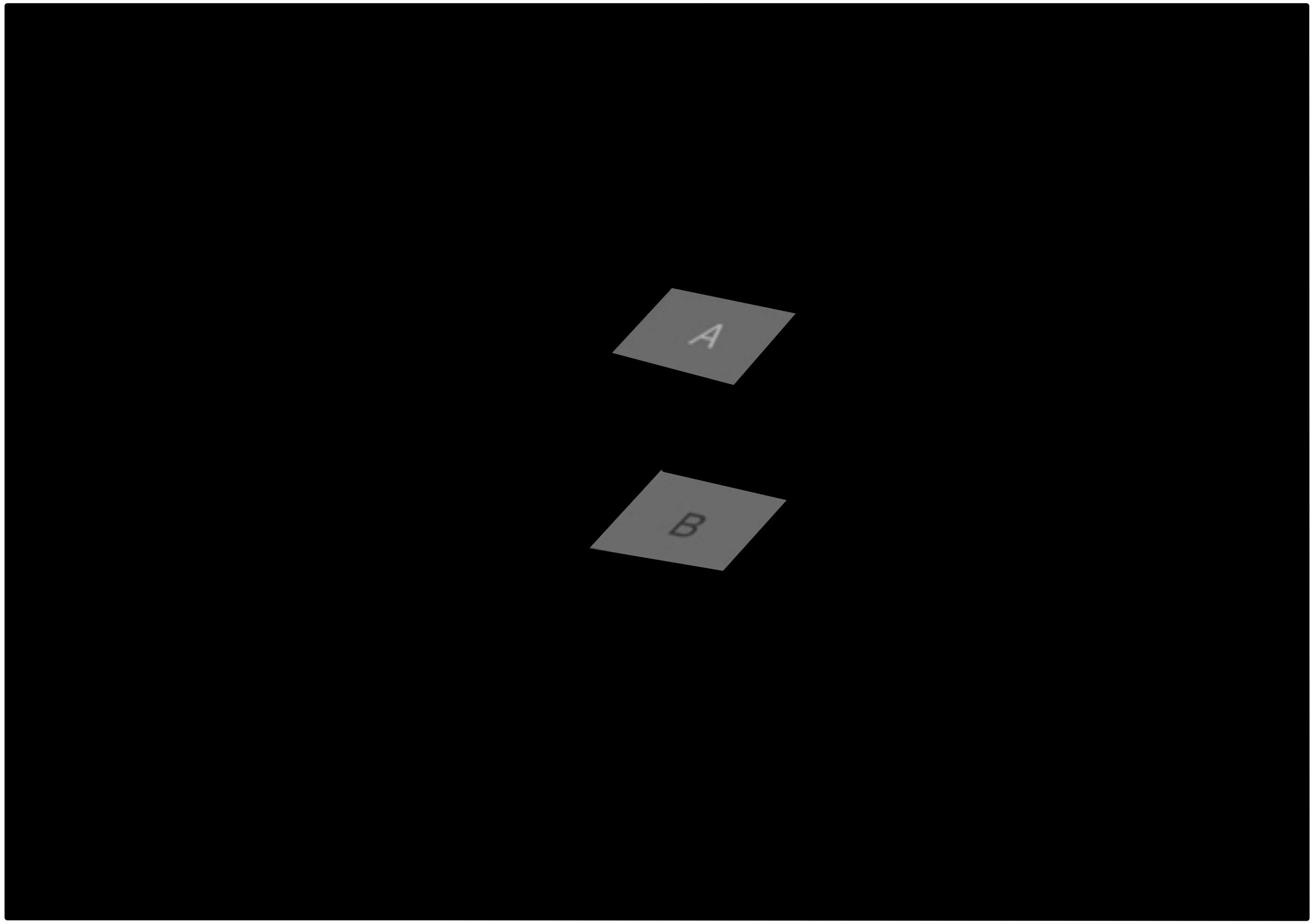
Depth manipulation



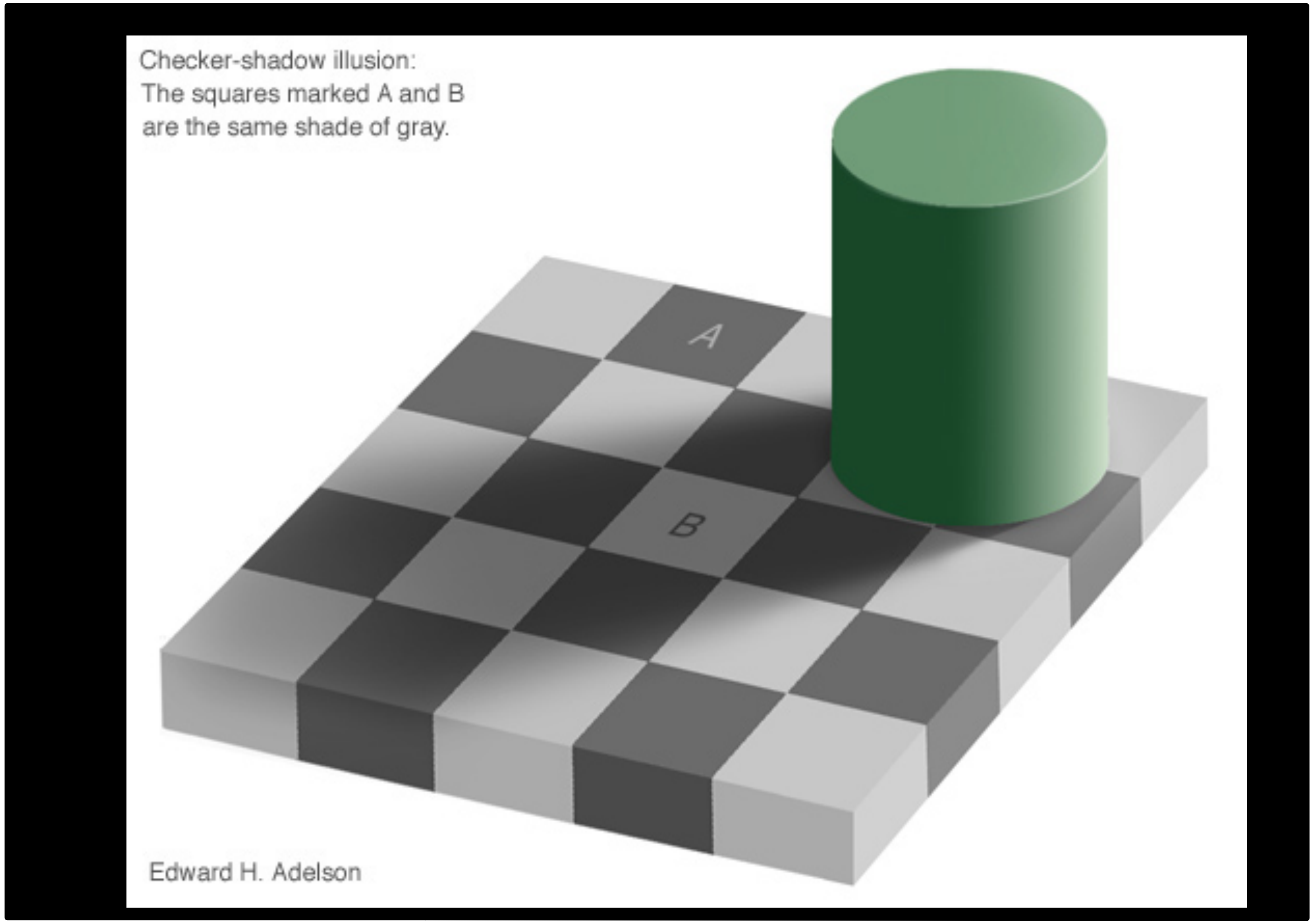
High(er) level vision

Simultaneous contrast

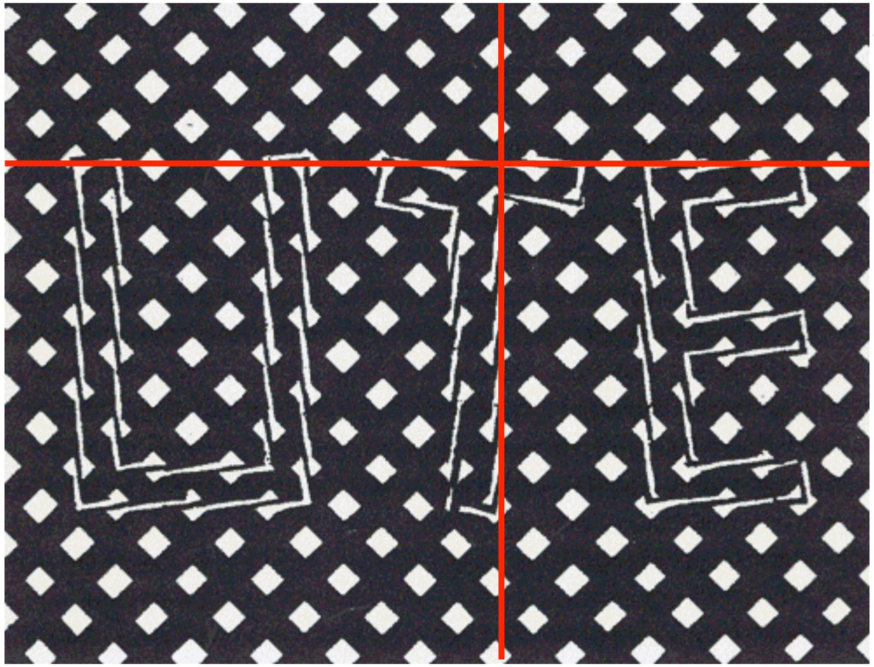
High-Level Contrast Processing

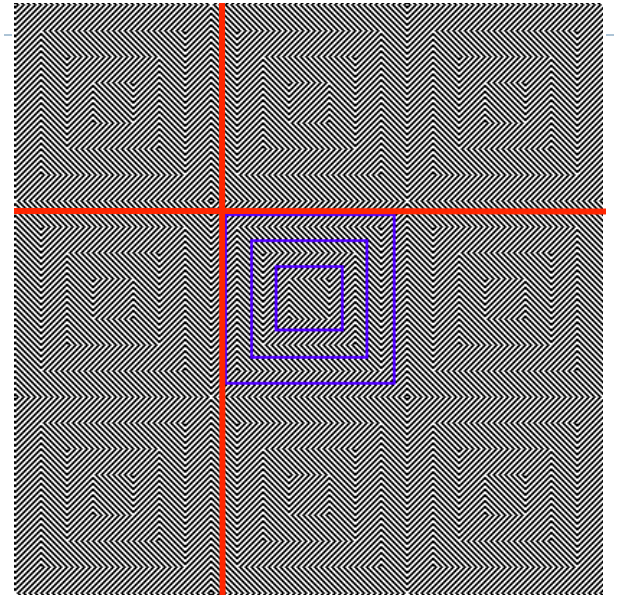


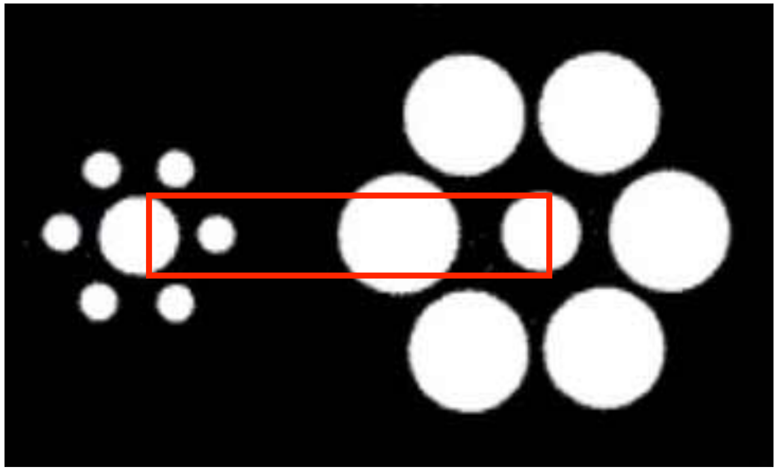
High-Level Contrast Processing



Shape Perception



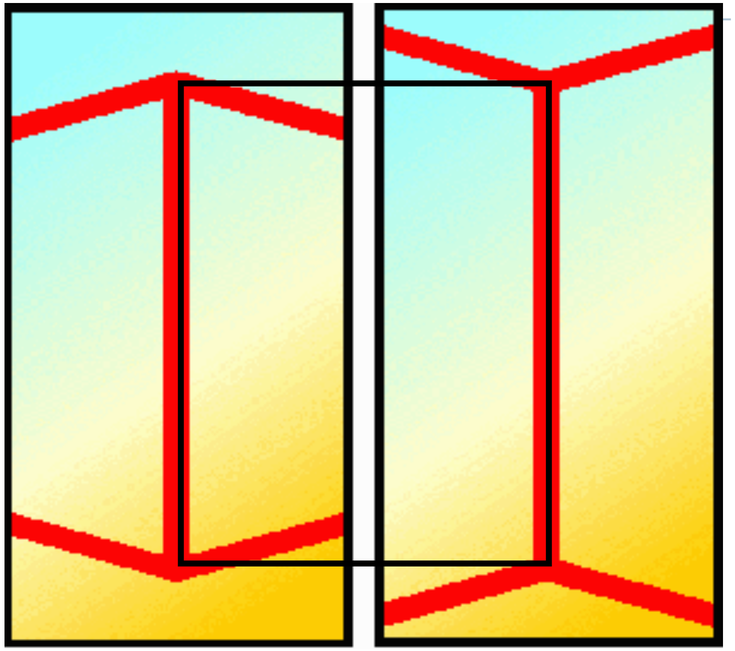


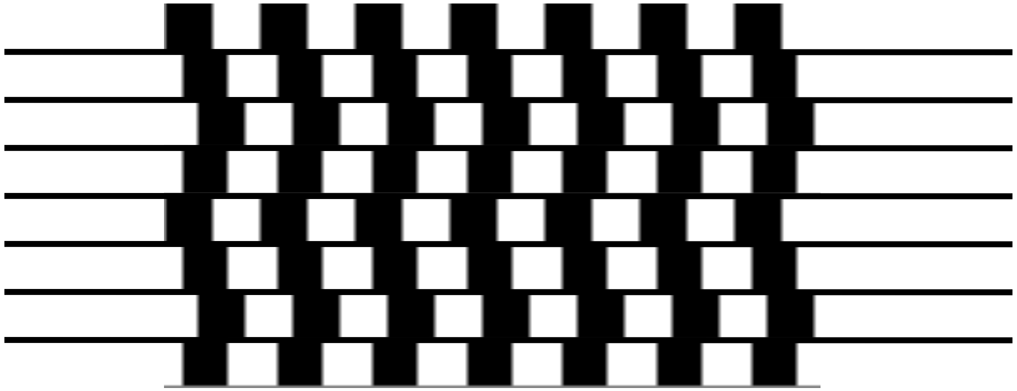


- Depends on surrounding primitives
  - Directional emphasis
  - Size emphasis

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

Shape Processing: Geometrical Clues



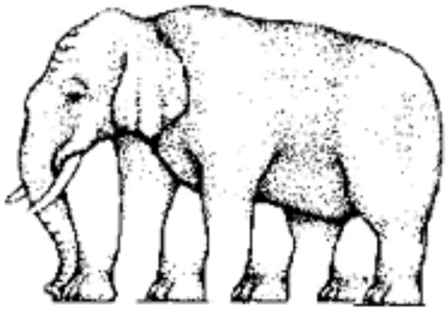


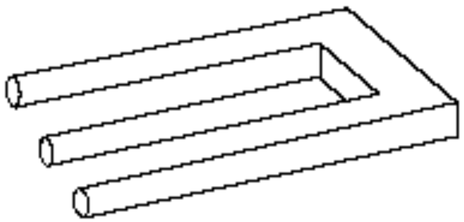
http://www.panoptikum.net/optischetaeuschungen/index.html

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

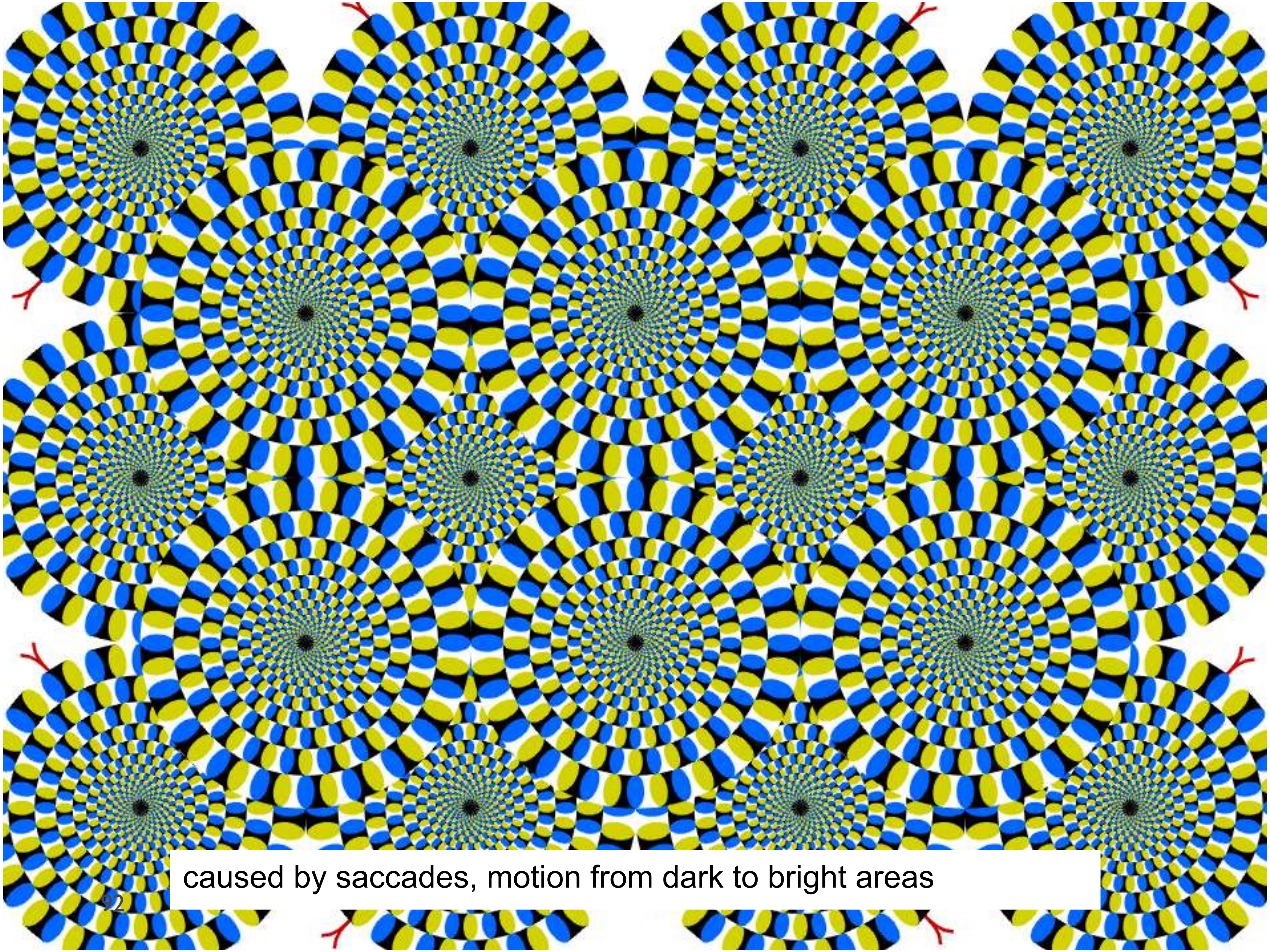
Impossible Scenes

- 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

Another Optical Illusion

- If you stare for approx. 20 seconds some of you will actually see a giraffe.

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