UNIVERSITY OF MICHIGAN

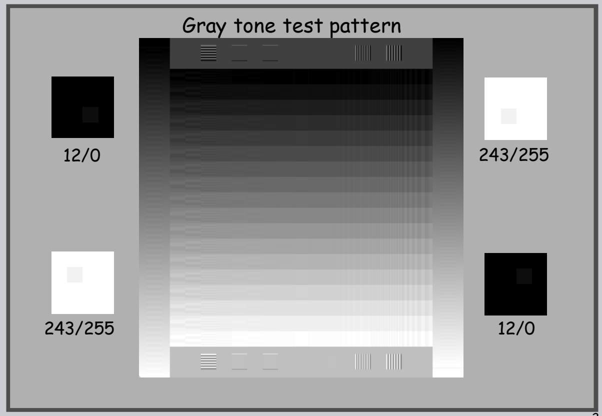
NERS/BIOE 481

Lecture 13
Observer Performance

Michael Flynn, Adjunct Prof
Nuclear Engr & Rad. Science
mikef@umich.edu
mikef@rad.hfh.edu

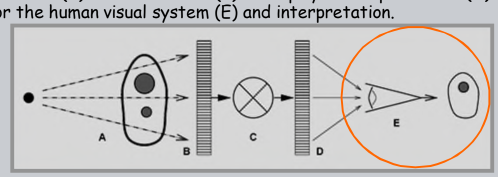
Henry Ford Health System
RADIOLOGY RESEARCH

Display Quality Test Image

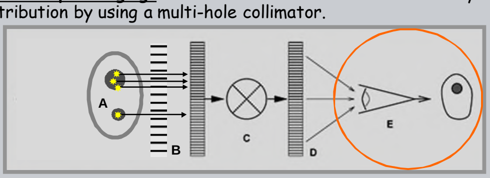


- General Models

Radiographic Imaging: Subject contrast (A) recorded by the detector (B) is transformed (C) to display values presented (D) for the human visual system (E) and interpretation.



Radioisotope Imaging: The detector records the radioactivity distribution by using a multi-hole collimator.

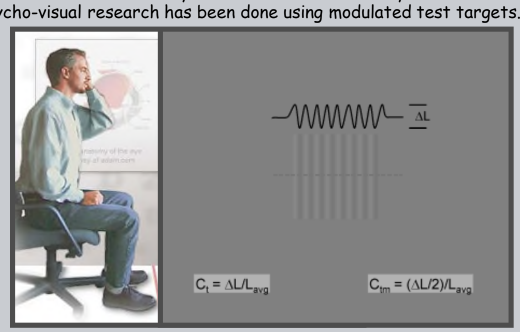


IX.A - Visual contrast threshold (15 charts)

A) Contrast Sensitivity of the Human Eye.
1) Test pattern characteristics
2) Contrast threshold/sensitivity
3) Measurement methods
4) Influence of size, frequency, & luminance
5) 2AFC measures of contrast sensitivity

IX.A.1 - Test patterns for visual performance

A variety of test patterns are used to assess visual performance. Clinical measures of acuity are done with a Snellen eye chart. Much psycho-visual research has been done using modulated test targets.



IX.A.2 - Contrast measures

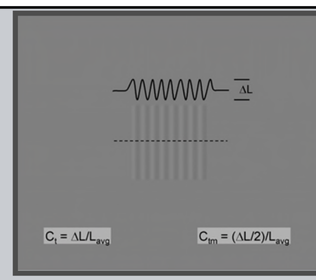
Contrast threshold: C_t , $C_{t\,m}$

The contrast for a just visible target.

Contrast sensitivity: C_s , C_{sm}

The inverse of the contrast threshold.

$C_s = 1/C_t$ $\quad C_{sm} = 1/C_{t\,m}$



Contrast is defined using two alternative definitions as illustrated.

- The early literature uses the Michelson definition of contrast threshold, C_{tm} , which is the amplitude of a sine function. This is used in Barten-1999.
- DICOM uses the peak to peak contrast, C_t , in part 14 of it's standard.

The Michelson contrast is one-half of the peak to peak contrast.

IX.A.3 - C_T Measurement Methods

Two methods to measure C_T

- Variable Adjustment
  - observer manipulates the contrast until C_T is found
  - dependent on the observer's confidence level
  - requires fine control of the contrast to find C_T
- Alternative Forced Choice (AFC)
  - observer must determine the location of the target from two (or more) options or make a guess.
  - does not require fine control of the contrast
  - dependent on a % correct criteria (for a 2AFC test, C_T = 75% chance of success)

IX.A.4 - Visual target characteristics.

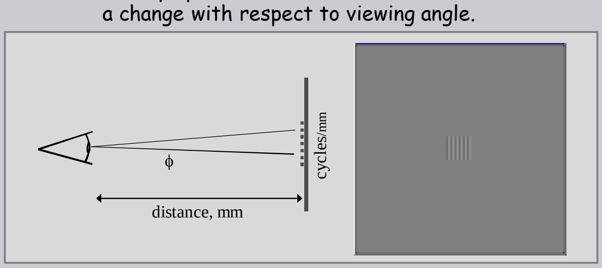
Barten fit a psycho-visual model function to the results of numerous experimental studies. In general, all studies used the variable adjustment method.

The following charts use Barten's model (Barten, SPIE, 1999) to illustrate how contrast threshold/sensitivity depends on the following characteristics of the target;

- Background Luminance
- Angular frequency,
- Target size
- Target orientation

IX.A.4 – Spatial Frequency: cycles/degree

The eye perceives luminance variations as a change with respect to viewing angle.

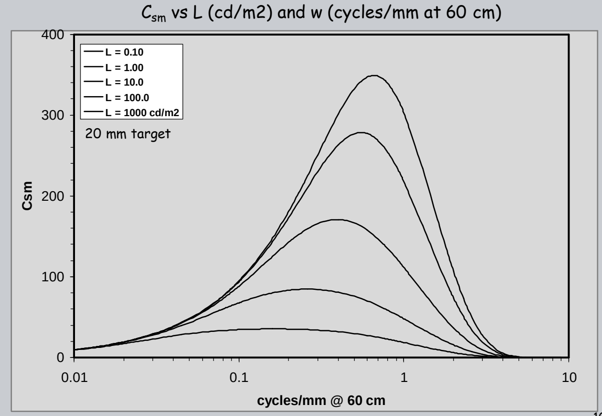


Data on visual performance can easily be converted from cycles/degree to cycles/mm at a specified viewing distance.

$$\text{cycles/mm} = \text{cycles/degree}\left(\frac{57.3}{\text{distance}}\right)$$

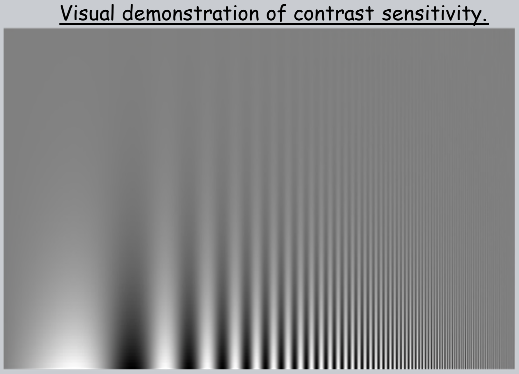
IX.A.4 - Contrast sensitivity vs luminance and frequency

C_{sm} vs L (cd/m2) and w (cycles/mm at 60 cm)



IX.A.4 - Contrast sensitivity vs luminance and frequency

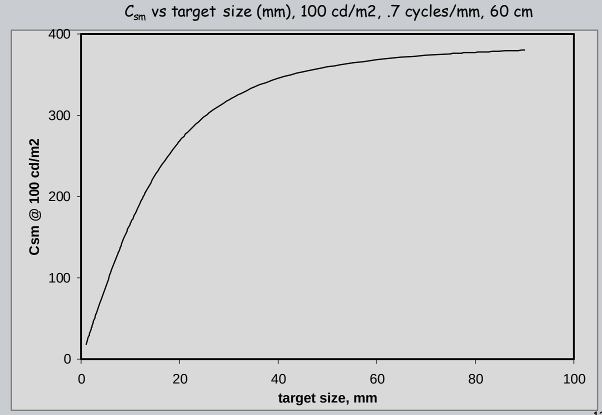
Visual demonstration of contrast sensitivity.



Campbell-Robson CSF chart

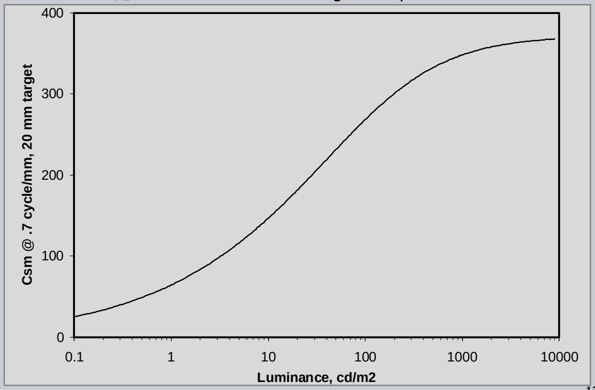
IX.A.4 - Contrast sensitivity vs target size

C_{sm} vs target size (mm), 100 cd/m2, .7 cycles/mm, 60 cm



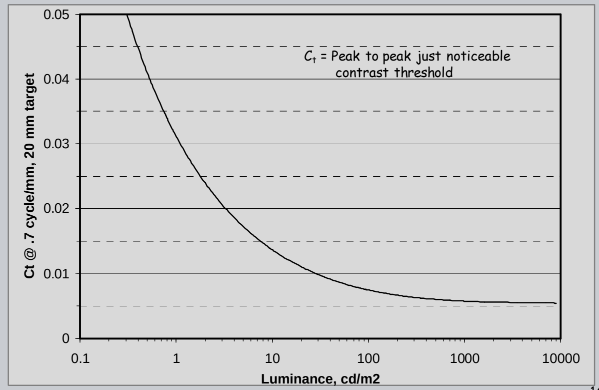
IX.A.4 - Contrast sensitivity vs luminance

C_{sm} vs L (cd/m2) , 20 mm target, .7 cycles/mm, 60 cm



IX.A.4 - Contrast threshold vs luminance

C_t vs L (cd/m2) , 20 mm target, .7 cycles/mm, 60 cm



IX.A.5 - Finding C_T for a 2AFC Observer Test

Two Alternative Forced Choice (2 AFC) method

- An observer views a series of image with a test pattern in one of 2 Alternative positions.
- For each, the observer makes a Forced Choice.

Data Analysis:

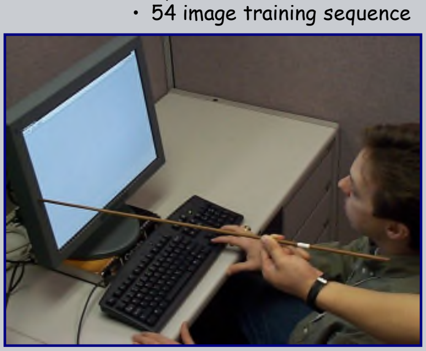
- Assume a model for the behavior of the human visual system (HVS)
- Identify the responses as (correct / incorrect) for images with varying contrast.
- Deduce contrast threshold (C_T = 75% correct) from a maximum likelihood fit of the HVS model

IX.A.5 - Graphics Software (2AFC test)

A series of bar patterns appear randomly in one of two regions. Observers must choose which side the target is on. Contrast varies randomly with each image

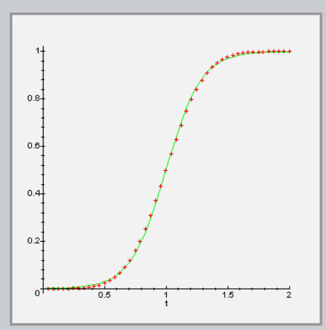
IX.A.5 - Display Conditions

- Minimal ambient luminance
- Observer level with target
- Eye 60 cm from monitor surface
- 54 image training sequence



IX.A.5 - The Psychometric Function

A psychometric expression is assumed for the probability that a grating target will be visually detected as a function of contrast.

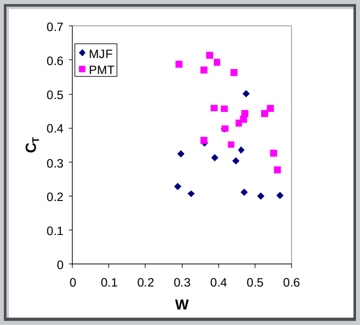


$$p(t) = 0.5\left[1 + \frac{1}{1 + e^{-2\left(\frac{t - C_T}{W}\right)}}\right]$$

IX.A.5 - Human C_T vs. W, two observers

Both C_T and **W** are determined from binary responses using maximum likelihood estimation (MLE).

- C_T is normalized here to be relative to the Barton model contrast threshold.
- C_T is referred to as a just noticeable difference (JND) unit.
- **W** is the width of the psychometric function in JND units.



For most person's C_T measured in a 2AFC experiment is less than that measured with the variable adjustment method.

IX.B - Human Vision & Display (25 charts)

Display requirements for the interpretation of radiological images are deduced from the performance of the **human visual system** (HVS).

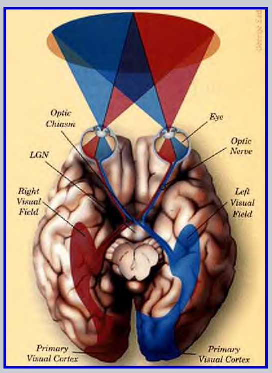
B) Human Vision & Display
1. Viewing Distance
2. Display Size
3. Pixel Size
4. Display Zoom
5. Equivalent Contrast

ACR–AAPM–SIIM TECHNICAL STANDARD FOR ELECTRONIC PRACTICE OF MEDICAL IMAGING
American College of Radiology, rev. 2017

IX.B.1 - Viewing Distance?

- Vergence
- Accomodation

- Vergence (convergence) allows both eyes to focus the object at the same place on the retina.
- The closer the object, the more the extraocular muscles converge the eyes inward towards the nose.



IX.B.1 – Viewing distance and vergence

Resting Point of Vergence

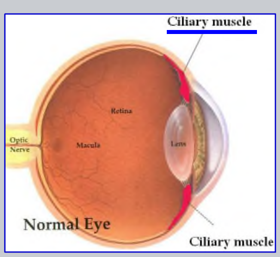
- Grandjean 1983
  - reported an average preferred viewing distance of 30 inches.
- Jaschcinsk-Kruza 1991
  - Objects closer than the resting point cause muscle strain.
  - The closer the distance, the greater the strain (Collins 1975).
- Jaschinski-Kruza 1998
  - Every one of the subjects studied judged an eye-screen distance of 20 inches to be too close.
  - All accepted a 40 inch distance.

→Arms length viewing distance: ~ 30 in

IX.B.1 – Viewing distance and accomodation

Resting Point of Accommodation

- The ciliary muscle changes the shape of the lens to focus at the distance of an object.
  - The eyes have a resting point of accommodation which is the distance that the eye focuses to when there is nothing to look at (Owens 1984).
  - This resting point averages about 31 inches (Krueger 1984).

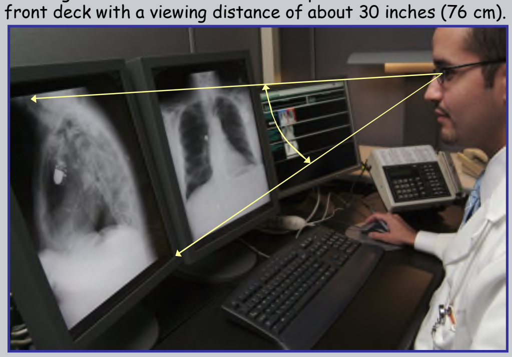


- Prolonged viewing of a monitor closer than the resting point of accommodation increases eye strain. The ciliary muscle must work 2.5 times harder to focus on a monitor 12 inches away than at 30 inches. (Jaschinski-Kruza 1988)

→Arms length viewing distance: ~ 30 in

IX.B.2 - Display Size?

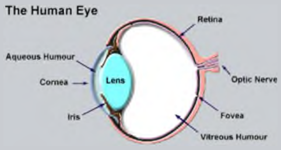
Radiologist at workstations with multiple monitors and a wide front deck with a viewing distance of about 30 inches (76 cm).



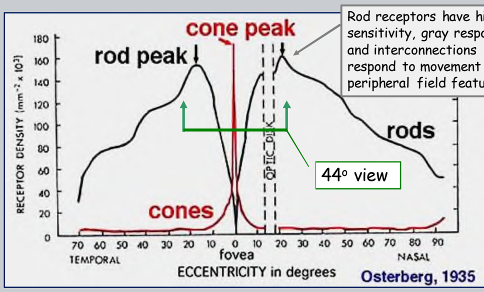
Angular field of view is measured using the diagonal distance.

IX.B.2 – HVS: peripheral response

The retina contains a large number of rod receptors (160 M) distributed over the peripheral field.



Rod receptors have high sensitivity, gray response, and interconnections that respond to movement of peripheral field features.



IX.B.2 – Display Size vs Viewing Distance

Visualization of the full scene is achieved when the diagonal display distance is about 80 % of the viewing distance.

- This corresponds to a viewing angle of 44 degrees.
- Somewhat larger than the peak retinal rod cell density

Task	Diagonal Size Inches (cm)	Viewing Distance Inches (cm)
Small Handheld	8 (20)	10 (25)
Tablet handheld	11 (28)	14 (36)
Laptop	16 (40)	20 (51)
Workstation	24 (61)	30 (76)

Note 1: The diagonal size of 22.5 inches for the workstation is similar to a traditional 14" x 17" radiographic film, 22.0"

Note 2: THX1 home entertainment recommends that the diagonal size should be about 84% of the viewing distance (46°).

IX.B.2 – Field of View

- 21 inch (diagonal) monitors with a field of 32 x 42 cm provide an effective size at a normal distance (30", 76 cm).
- 30 inch (diagonal) wide format (16:9) monitors provide effective image size when split into two frames of 20" size.

Eizo GX1030
30" diagonal, 4096 x 2560, 0.158 mm pitch

Eizo GX540 dual
21" diagonal, 2048 x 2560, 0.165 mm pitch

IX.B.3 – Pixel Size?

Pixel pitch:

"For monitors used in diagnostic interpretation, it is recommended that the pixel pitch be about 0.200 mm and not larger than 0.210 mm."

"For this pixel pitch, individual pixels and their substructure are not visible and images have continuous tone appearance."

"No advantage is derived from using a smaller pixel pitch since higher spatial frequencies are not perceived."

American College of Radiology (ACR) Guidelines.

Retina Display is a brand name used by Apple for liquid crystal displays that, according to Apple, have a high enough pixel density that the human eye is unable to notice pixelation at a typical viewing distance. (http://en.wikipedia.org/wiki/Retina_Display)

IX.B.3 – HVS: Retinal anatomy

The spacing of cells in the retina of the human eye limit the maximum spatial frequency (cycles/degree)

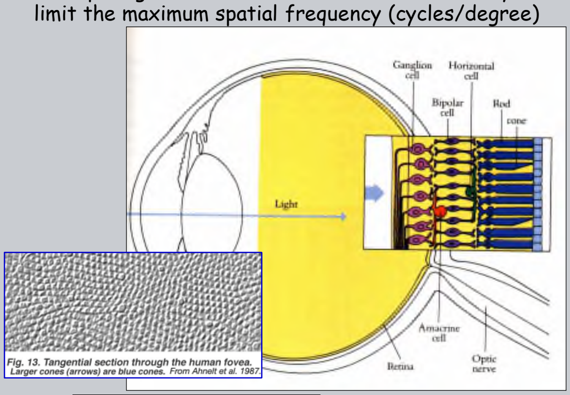
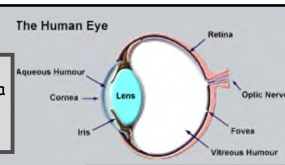


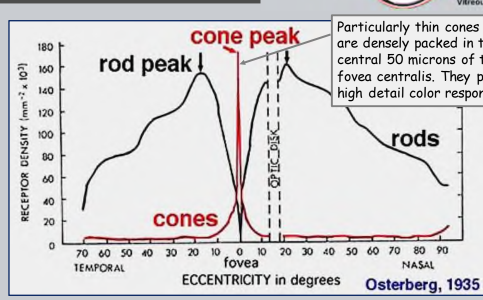
Fig. 13. Tangential section through the human fovea. Larger cones (arrows) are blue cones. From Ahnelt et al. 1987

IX.B.3 – HVS: Foveal response

At 60 cm, 1 degree corresponds to a 1 cm field of view. This area is focused on a 288 micron region of the retina, the fovea.

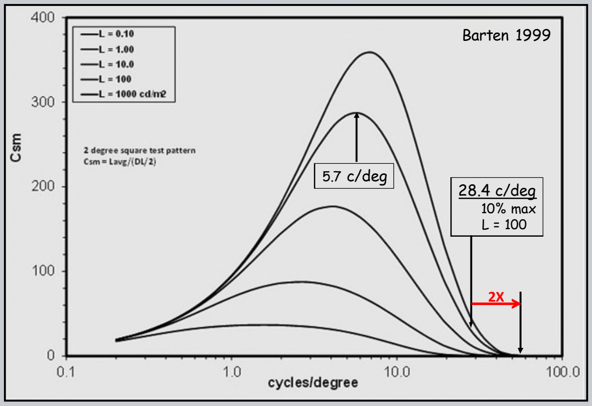


Particularly thin cones (2 μm) are densely packed in the central 50 microns of the fovea centralis. They provide high detail color response.



IX.B.3 – Contrast Sensitivity as a measure of spatial acuity

See slide 10



Note: Contrast sensitivity is the inverse of contrast threshold

IX.B.3 – Pixel Size at Maximum Spatial Acuity

- The visual spatial frequency limit and associated pixel size can be defined as that for which Cs = 10% of maximum (100 cd/m2).
- The pixel size of a display system that matches the resolving power of the human eye depends on the observation distance.

	View Distance Inches (cm)	Diagonal Size Inches (cm)	Pixel Pitch mm	Pixels per inch PPI
Small Handheld	10 (25)	8 (20)	78	325
Tablet handheld	14 (36)	11 (28)	109	232
Laptop	20 (51)	16 (40)	156	163
Workstation	30 (76)	24 (61)	234	108

- Two pixels per cycle are assumed based on the Nyquist theorem.
- No pixel structure artifacts are noticeable for these pixel sizes.
- No advantage is gained by using smaller pixel sizes.

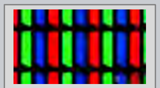
$P_P = D_V / 3255$ => 3255 = 2 x 57.3 x 28.4

$P_P = 0.307\ D_V$ => D_V in meter & P_P in mm

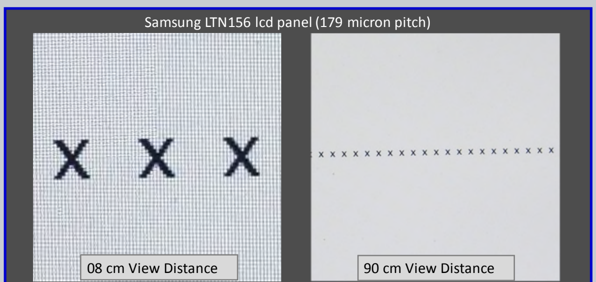
Note: values are consistent with Apple retinal display.

IX.B.3 – Pixel Size at Maximum Spatial Acuity

For pixel pitches that are too large for the viewing distance used, pixel structure details appear as a textured pattern.



LTN pixel structure

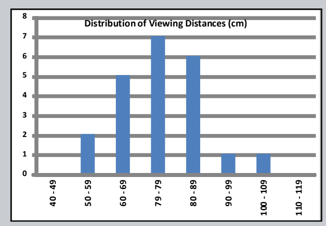


Illustrated appearance of X pattern at two viewing distances.

IX.B.3 – Pixel Size at Maximum Spatial Acuity

- The ACR recommended pitch of 0.200 mm results in continuous tone display (i.e. no visible pixel structure) for viewing distances larger than 65 cm.
- At HFHS, most radiologist read at a distance slightly larger than 65 cm.

- 22 Staff Radiologists
- Mean: 76.7 cm
- STD: 11.4 cm
- Range: 65 to 88 cm
- 19 of 22 were equal or greater than 65 cm.



$P_P = 0.307\ D_V$, for D_V in meter & P_P in mm

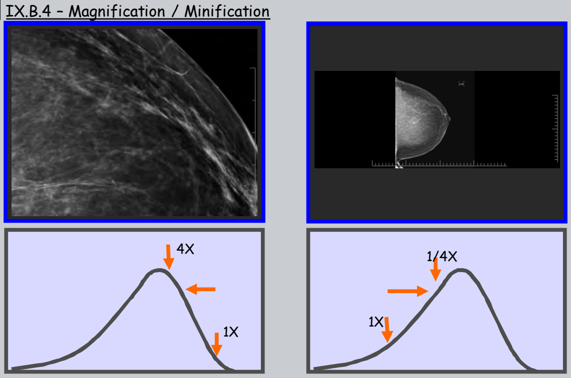
IX.B.4 – Display Zoom?

Detector Detail
in relation to
Display Acuity

IX.B.4 – Viewing distance and image zoom

- Use of image zoom features is ergonomically better than leaning forward for close inspection.
- Split deck tables with a broad front deck usefully prohibit close inspection with 3 MP monitors.

IX.B.4 – Magnification / Minification



Magnification is used to display detail at the detector pixel level with good contrast sensitivity.

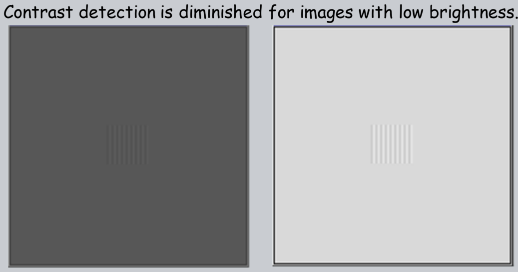
Minification is used to increase the spatial frequency of diffuse structures.

IX.B.5 – Equivalent Contrast?

- Grayscale response
- Luminance ratio (L'max/L'min)

IX.B.5 – Contrast detection in relation to brightness

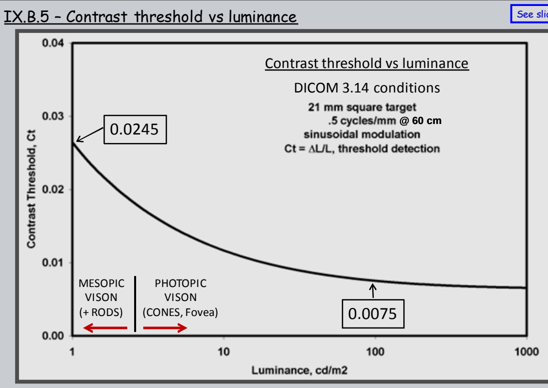
- Contrast detection is diminished for images with low brightness.



- Extensive experimental models have documented the dependence of contrast detection on luminance, spatial frequency, orientation and other factors. The empirical models of either S. Daly or J. Barton provide useful descriptions of this experimental data.

IX.B.5 – Contrast threshold vs luminance

See slide 19

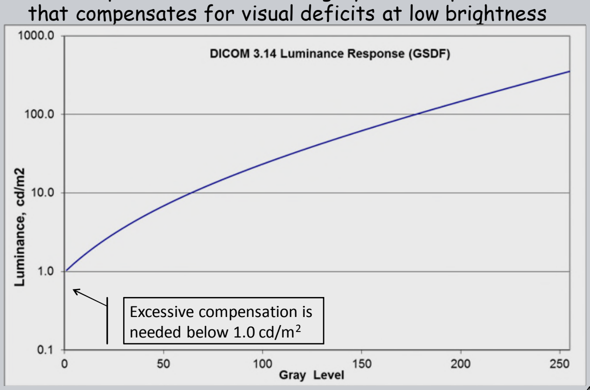


The Barton model describes the average contrast threshold of normal observers. Significant differences exist for individual observers for different test methods

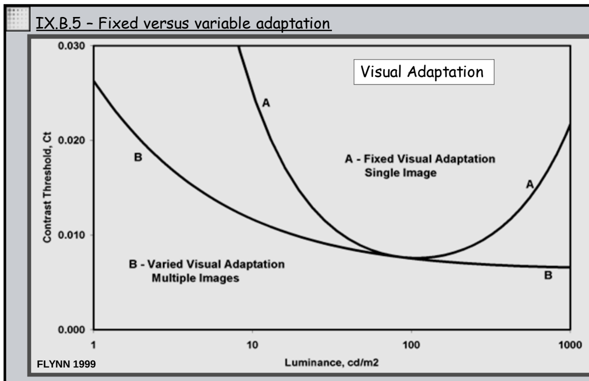
IX.B.5 – DICOM graylscale display standard

See Lecture 12 (VIII.C.b.2)

DICOM part 3.14 describes a grayscale response that compensates for visual deficits at low brightness



IX.B.5 – Fixed versus variable adaptation

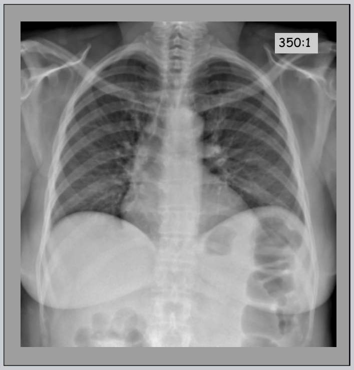


The contrast threshold, ΔL/L, for a just noticeable difference (JND) depends on whether the observer has fixed (B) or varied (A) adaptation to the light and dark regions of an overall scene.

IX.B.5 – Effect of Lmax/Lmin

- Medical images should be displayed using a luminance range of about
  350:1.
- Images prepared for range of 350 that are display on a monitor with large range will have poorly perceived contrast in dark regions.

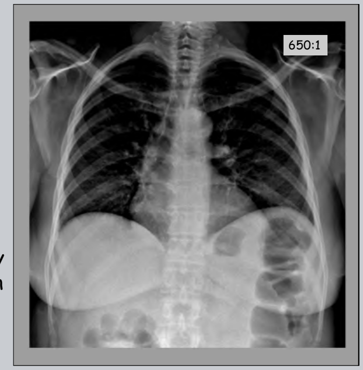
350:1 → .1 to 2.65 OD
650:1 → .1 to 2.90 OD



IX.B.5 – Effect of Lmax/Lmin

- Medical images should be displayed using a luminance range of about
  350:1.
- Images prepared for range of 350 that are display on a monitor with large range will have poorly perceived contrast in dark regions.

350:1 → .1 to 2.65 OD
650:1 → .1 to 2.90 OD



IX.B – Display Specifications, Summary

Summary
Recommended Luminance Response Specifications

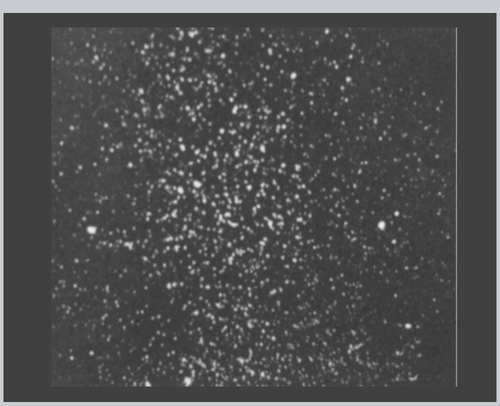
	Diagnostic	Other
L_{min}:	≥ 1.0 cd/m^2	≥ 0.8 cd/m^2
L_{max}:	≥ 350 cd/m^2	≥ 250 cd/m^2
Luminance ratio (LR)	~350 (≥ 250).	~350 (≥ 250).
Luminance response	GSDF	GSDF
GSDF tolerance	10%	20%
Pixel pitch mm	210 mm	~250 (<300)

- L_{amb} less than 1/4th of L_{min}.
- Diagonal size of 20-24 inches with 3:4 or 4:5 aspect
- D65 (6500 C) white point

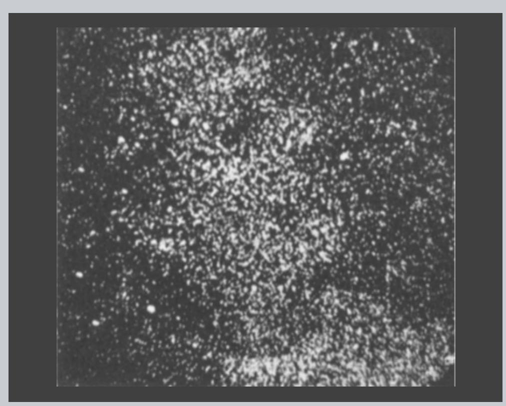
IX.C – Detection of targets in noise (12 charts)

C) Detection of targets in noise
1) Image noise & the Rose model
2) Complex noise patterns

C.1 - Noise & Quantum Mottle



C.1 - Noise & Quantum Mottle



C.1 - Noise & Quantum Mottle

C.1 - Noise & Quantum Mottle

C.1 - Noise & Quantum Mottle

C.1 - Noise & Quantum Mottle

Illustrations from;
Rose A, Vision – Human and Electronic, Plenum Press

C.2 - Signal to Noise Ratio

For photon imaging:

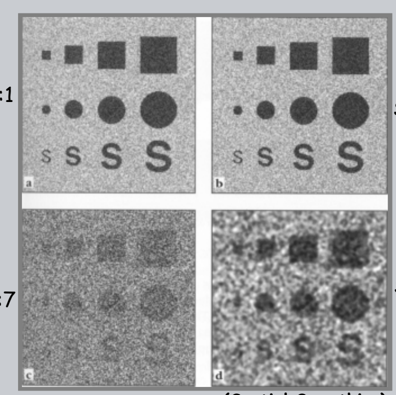
- Signal Proportional to number of photons, Q
- Noise Approximated by standard deviation, σ
- Standard Deviation Equals Square root of Q (Poisson Statistics)

$$\frac{\text{Signal}}{\text{Noise}} = \frac{Q}{\sigma} = \frac{Q}{\sqrt{Q}} = \sqrt{Q}$$

C.2 - Signal to Noise Ratio

SNR 1:1

SNR 1:3



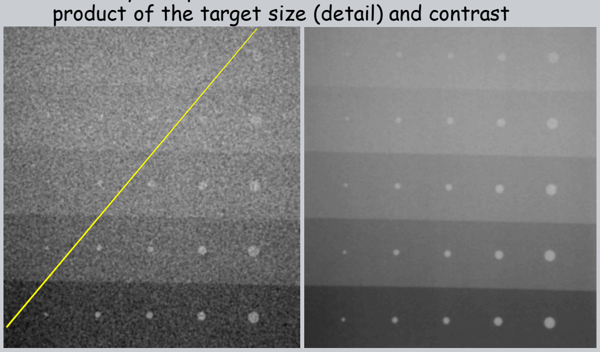
SNR 1:7

SNR 1:7

(Spatial Smoothing)

C.2 - Contrast Detail & noise

Visibility at a particular SNR is related to the product of the target size (detail) and contrast



Fluoroscopy (0.74 μR/fr) SNR low — Radiography (353 μR/fr) SNR high

C.2 - The Rose model.

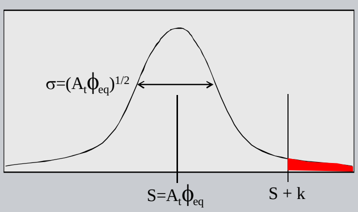
- The ability of an observer to detect a low contrast target in a uniform background can be modeled by considering the background noise for regions equal to the target area in relation to the absolute contrast of the target.
- This can be estimated by considered the product of the target area, A_{tar}, and the noise equivalent quanta, ϕ_{eq}, and using the relative contrast to convert the signal to noise ratio to the contrast to noise ratio

$$\frac{\text{Signal}}{\text{Noise}} = \frac{S}{N} = \left(A_{tar}\ \varphi_{eq}\right)^{1/2}$$

$$\frac{\text{Contrast}}{\text{Noise}} = C_r\frac{S}{N} = C_r\left(A_{tar}\ \varphi_{eq}\right)^{1/2}$$

C.2 - The Rose contrast-area relationship.

- A criteria for the detection of a target with specified contrast is that there be no regions in the background with area equal to the target area for which the average image signal variation from random noise is equal to or greater than the target contrast.
- The random distribution of signal values from many areas in the background is described by gaussian probablility distribution function.



k	Prob S > S+k
1σ	0.15
2σ	0.023
3σ	1.3 x 10^{-3}
4σ	3 x 10^{-5}
5σ	3 x 10^{-7}
6σ	2 x 10^{-9}

C.2 - The Rose model.

- The background region may have a large number of regions that may cause a false impression of a target. The criteria for detection should thus be 4-5 times the background standard deviation.
- We thus require that the contrast to noise ratio be larger than a threshold value (k_t) of 4-5 for a target object to be detected on a uniform background of noise.
- The minimum, or threshold, relative contrast for a target to be detected can thus be written as

$$k_t \le \left|\text{Contrast}/\text{Noise}\right| \qquad k_t \sim 4\text{-}5$$

$$k_t \le \left|C_r\right|\left(A_{tar}\varphi_{eq}\right)^{1/2}$$

$$k_t^2 \le C_r^2 A_{tar}\varphi_{eq}$$

See Rose, pg 26

IX.D – Statistical Performance of Observers (16 charts)

D) Statistical Performance of Observers

1) Sensitivity / Specificity
2) Predictive value
3) The ROC curve
4) Agreement & Kappa
5) Attention Effect

D.1 - Interpretations in relation to Findings

When radiologic examinations are interpreted to determine the presence or absence of a finding of interest, 4 scenarios are possible;

- True Positive (TP),
  The finding is PRESENT and was IDENTIFIED.
- False Negative (FN),
  The finding is PRESENT but was NOT IDENTIFIED .
- False Positive (FP),
  The finding is NOT PRESENT but was IDENTIFIED.
- True Negative (TN),
  The finding is NOT PRESENT and was NOT IDENTIFIED.

The term ‘finding’ is used here to indicate a particular image feature that may be indicative of a disease (a nodule associated with cancer) or condition (a fracture).

D.1 - Sensitivity and Specificity

Consider an experiment in which 100 cases with a finding of interest and 100 cases without the finding are presented for interpretation.

Interpretation	Finding: Present		Finding: Absent	
Positive	TP	90	FP	10
Negative	FN	10	TN	90
Total=200		100		100

Sen = 90%

Spe = 90%

- Sensitivity:
  Fraction of cases with the finding that were correctly interpreted as positive. $Sensitivity = \frac{TP}{TP + FN}$
- Specificity:
  Fraction of cases without the finding that were correctly interpreted as negative. $Specificity = \frac{TN}{TN + FP}$

D.2 - Predictive Value

In practice, as opposed to experiment, the fraction of all cases having findings present is defined as the prevalence, P.

Sensitivity 90% , Specificity 90% , Prevalence 1/11

Interpretation	Present		Absent		Predictive Value	
Positive	TP	90	FP	100	PPV	90/190 = .474
Negative	FN	10	TN	900	NPV	900/910 = .989
Total=1100	Tot x P = 100		Tot x (1-P) =1000			

- Positive Predictive Value:
  Fraction of positive interpretations that have findings present. $PPV = \frac{TP}{TP + FP}$
- Negative Predictive Value:
  Fraction of negative interpretations that do not have findings present. $NPV = \frac{TN}{TN + FN}$

D.2 - Predictive Value

From the definition of sensitivity and specificity, we can deduce TP and TN as a function of prevalence..

$$Sensitivity = Sen = \frac{TP}{TP + FN} = \frac{TP}{Total \times P_r}$$

$$Specificity = Spe = \frac{TN}{TN + FP} = \frac{TN}{Total \times (1 - P_r)}$$

=>

$$TP = Sen\left(Total \times P_r\right)$$

$$TN = Spe\left[Total \times (1 - P_r)\right]$$

We then note that;

$$FP = \left[Total \times (1 - P_r)\right] - TN = (1 - Spe)\left[Total \times (1 - P_r)\right]$$

Thus;

$$PPV = \frac{TP}{TP + FP} = \frac{Sen \times P_r}{(Sen \times P_r) + (1 - Spe)(1 - P_r)}$$

And similarly;

$$NPV = \frac{TN}{TN + FN} = \frac{Spe \times (1 - P_r)}{((1 - Sen) \times P_r) + Spe \times (1 - P_r)}$$

D.2 - Predictive Value and Prevalence

The prevalence influences the PPV and NPV

Sensitivity 90% , Specificity 90% , Prevalence 1/101

Interpretation	Present		Absent		Predictive Value	
Positive	TP	90	FP	1000	PPV	90/1090 = .083
Negative	FN	10	TN	9,000	NPV	9000/9010 = .999
Total=10100	T x P = 100		Tx(1-P)=10,000			

- Positive Predictive Value:
  Fraction of positive interpretations that have findings present. $PPV = \frac{TP}{TP + FP}$
- Negative Predictive Value:
  Fraction of negative interpretations that do not have findings present. $NPV = \frac{TN}{TN + FN}$

D.2 - Predictive Value and Prevalence

The prevalence influences the PPV and NPV

Sensitivity 80% ,Specificity 96% ,Prevalence 1/101

Interpretation	Present		Absent		Predictive Value	
Positive	TP	80	FP	400	PPV	80/480 = .167
Negative	FN	20	TN	9,600	NPV	9600/9620 = .998
Total=10100	T x P = 100		Tx(1-P)=10,000			

Interpreting exams 'cautiously' such that only a definite finding is read as positive;

- Reduces the sensitivity
- Increases the specificity
- and changes the predictive values.

Kavanagh 2000
J. Med. Screen

Sensitivity:	76%
Specificity:	95%
Prevalence:	.007
PPV:	9.2%

96420 patients.

D.2 - Important concepts

- Sensitivity and specificity are determined from experiments where the findings are known by independent methods ('gold standards').
- Predictive value is determined from the prevalence of the finding in the clinical population and measured values of specificity and sensitivity.

D.3 Receiver Operating Characteristics (ROC)

- 'cautious' interpretation such that only a definite finding is read as positive results in

  high sensitivity and low specificity

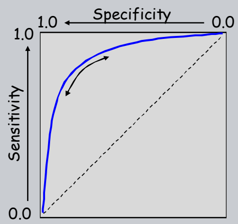
- 'aggressive' interpretation such that the suggestion of a finding is read as positive results in

  low sensitivity and high specificity.

- Varying the criteria for interpreting findings results in a range of (sensitivity, specificity) combinations.

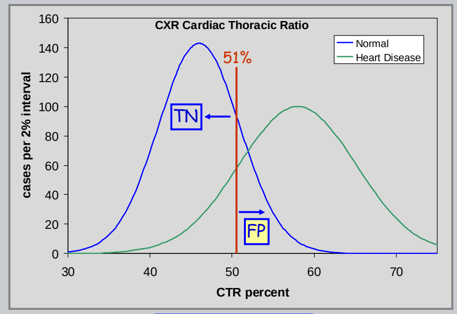
- The operating characteristics of an interpreter (receiver) are described by plotting sensitivity vs specificity.
- This is the ROC curve.

Peterson WW, Birdsall TG,
The Theory of Signal Detectability
TR 13, EE dept, Univ of MI, 1953



D.3 – distribution of responses

Turner illustrates sensitivity and specificity using the cardiac thoracic ratio observed from chest x-rays as an indicator of heart disease.

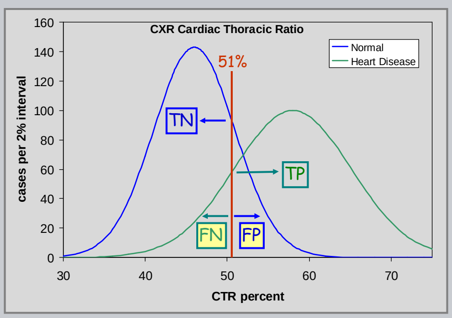


51% criteria

TN = 752 , FP = 139
Specificity = 0.84

D.3 – decision criteria, 51%

A decision criteria establishes a percent ratio above which the finding is interpreted as abnormal. At 51% Sensitivity = Specificity = 0.84 .



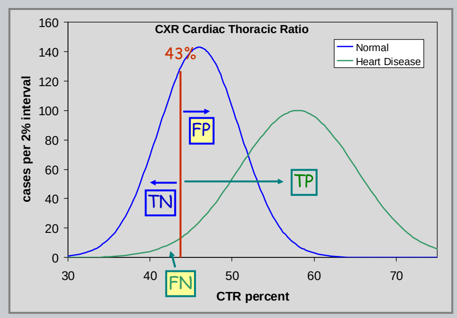
51% criteria

TN = 752 , FP = 139
Specificity = 0.84

TP = 745 , FN = 143
Sensitivity = 0.84

D.3 – decision criteria, 43%

Reducing the criteria to 43% results in a very good sensitivity.



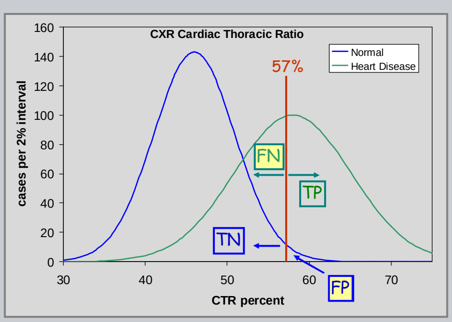
43% criteria

TN = 242 , FP = 649
Specificity = 0.27

TP = 843 , FN = 15
Sensitivity = 0.98

D.3 – decision criteria, 57%

Increasing the criteria to 57% results in a very good sensitivity.



57% criteria

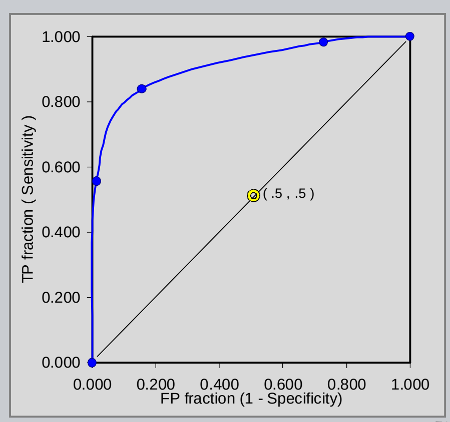
TN = 879 , FP = 12
Specificity = 0.99

TP = 494 , FN = 394
Sensitivity = 0.56

D.3 – ROC curve

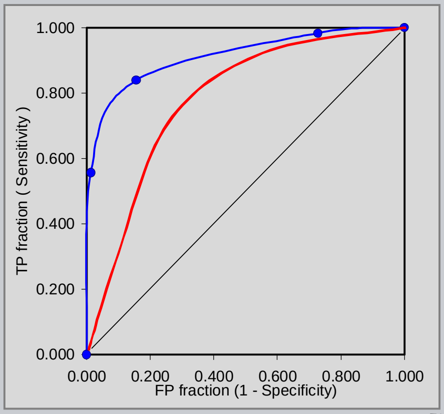
These 3 values of (Sens,1-Spec) along with the limiting values of (0,0) and (1,1) describe the ROC for this test.

If images are randomly found as positive or negative without looking at them, the response is along the diagnonal line.



D.3 – ROC curve area

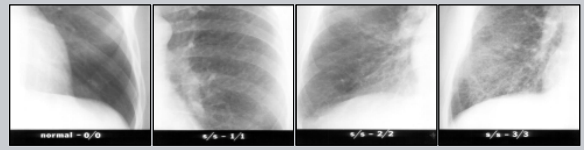
The area under ROC curves can be used as a measure of whether one test is better than another.



D.4 – Agreement and the Kappa statistic

Radiation images are sometimes evaluated using a grading scale for the appearance of specific image characteristics.

An example is the classification of pneumoconiosis using a scale developed by the International Labor Office (ILO) to describe small opacities observed in lung radiographs.



This has been used worldwide to evaluate occupational diseases in workers exposed to excessive dust (coal miners ...)

D.4 – Agreement and the Kappa statistic

Halldin 2014 reported on the agreement between classifications with done using new digital radiography reference standards (DR) and done with the traditional film reference standards.

For this reader, the Kappa statistic, K, indicates moderate agreement

$DR_{ILO2011\text{-}D}$

FSR	0/0	0/1	1/0	1/1	1/2	2/1	2/2	2/3	3/2	3/3	Total
0/0	6	6	8	2							22
0/1	6	7	18	6							37
1/0	5	13	35	17	4	3					77
1/1		1	6	5	2		1				15
1/2					1	2	1				4
2/1			2		1	2	1				6
2/2			1		1		1	1			4
2/3							2	1			3
3/2											0
3/3								1			1
Total	17	27	70	30	9	7	6	3	0	0	169

K= 0.40 (0.30-0.51)

Halldin et.al., Validation of the International Labour Office Digitized Standard Images for Recognition and Classification of Radiographs of Pneumoconiosis, Academic Radiology, Mar., 2014.

D.4 – Agreement and the Kappa statistic

- Cohen's kappa measures the agreement between two raters.
- Weighted kappa lets you count disagreements differently and is useful when codes are ordered.

$$\kappa = 1 - \frac{\left(1 - \sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}x_{ij}\right)}{\left(1 - \sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}m_{ij}\right)}$$

Values of K	agreement
< 0.20	Poor
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Good
0.81 - 1.00	Very good

w_{ij} matrix of weighting values

x_{ij} matrix of observed scores

m_{ij} expected scores (chance distribution)

Cohen, J. (1968). "Weighed kappa: Nominal scale agreement with provision for scaled disagreement or partial credit". Psychological Bulletin 70 (4): 213–220
http://en.wikipedia.org/wiki/Cohen%27s_kappa

D.4 – Agreement and the Kappa statistic

$$\kappa = 1 - \frac{\left(1 - \sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}x_{ij}\right)}{\left(1 - \sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}m_{ij}\right)}$$

Example matrices: Weighted Kappa = .55

Observed

j \ i	1	2	3	4	5	
1	27	10	4	2	1	44
2	10	18	10	4	2	44
3	4	10	16	10	4	44
4	2	4	10	18	10	44
5	1	2	4	10	27	44
	44	44	44	44	44	220

Expected (chance)

j \ i		1	2	3	4	5	
1	1	8.8	8.8	8.8	8.8	8.8	44
2	2	8.8	8.8	8.8	8.8	8.8	44
3	3	8.8	8.8	8.8	8.8	8.8	44
4	4	8.8	8.8	8.8	8.8	8.8	44
5	5	8.8	8.8	8.8	8.8	8.8	44
		44	44	44	44	44	220

- The observed matrix of scores was hypothetically filled to give equal probablility distributions for both observers, i and j.
- Thus, the expected matrix has equal values.
- A Kappa of .55 is computed for a weights which are linear with distance from the diagonal.

Linear Weight

0.00	0.25	0.50	0.75	1.00
0.25	0.00	0.25	0.50	0.75
0.50	0.25	0.00	0.25	0.50
0.75	0.50	0.25	0.00	0.25
1.00	0.75	0.50	0.25	0.00

D.5 - Selective Attention

Selective Attention

Daniel J. Simons

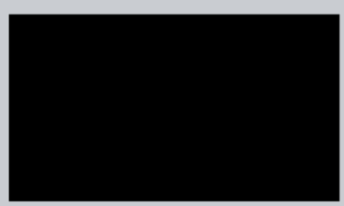
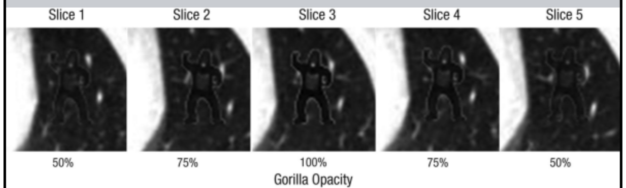


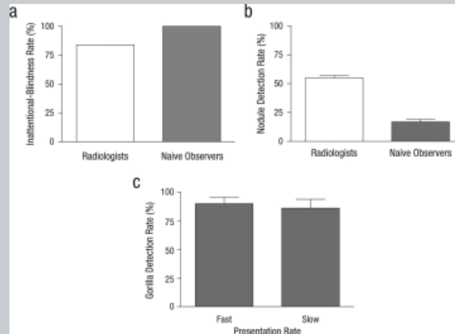
Fig. 1. Illustration of the slices showing the gorilla in the final trial of Experiments 1 and 2.



•Drew T et al. Psychological Science 2013;24:1848-1853

Psychological SCIENCE

Fig. 3. Experimental results.



•Drew T et al. Psychological Science 2013;24:1848-1853

Psychological SCIENCE