Chapter 18: Image Perception and Assessment


*Diagnostic Radiology Physics: A Handbook for Teachers and Students*

Objective:
To provide the student with an introduction to human visual perception and to task-based objective assessment of an imaging system.

Slide set prepared by E. Berry (Leeds, UK and The Open University in London)
# CHAPTER 18

## TABLE OF CONTENTS

18.1. Introduction  
18.2. The Human Visual System  
18.3. Specifications of Observer Performance  
18.4. Experimental Methodologies  
18.5. Observer Models
18.1 INTRODUCTION
Introduction (1 of 2)

- The main purpose of a medical image is to provide information to a human reader, such as a radiologist, so that a diagnosis can be reached - rather than to display the beauty of the human internal workings.

- It is important to understand how the human visual system affects the perception of contrast and spatial resolution of structures that are present in the image.

- If the image is not properly displayed, or the environment is not appropriate, subtle clinical signs may go unnoticed, which can potentially lead to a misdiagnosis.
Introduction (2 of 2)

- This chapter provides an introduction to human visual perception and to task-based objective assessment of an imaging system.

- Model for the contrast sensitivity of the human visual system:
  - used to derive the gray-scale standard display function (GSDF) for medical displays.

- Task-based assessment measures the quality of an imaging system:
  - requires a measure of the ability of an observer to perform a well-defined task, based on a set of images.
  - metrics for observer performance.
  - experimental methodologies for the measurement of human performance.
  - estimation of task performance based on mathematical observer models.
18.2 THE HUMAN VISUAL SYSTEM
The human visual system

- 18.2.1 The human eye
- 18.2.2 The Barten model
- 18.2.3 Perceptual linearization
- 18.2.4 Viewing conditions
18.2 THE HUMAN VISUAL SYSTEM

18.2.1 THE HUMAN EYE
18.2 THE HUMAN VISUAL SYSTEM

18.2 The human visual system

- 18.2.1 The human eye
- 18.2.2 The Barten model
- 18.2.3 Perceptual linearization
- 18.2.4 Viewing conditions
Structure of the human eye

- The human eye is a complex organ that processes visual images and relates information to the brain (Dragoi).
- The retina is the light-sensitive part of the eye.
- There are two types of photosensitive cells in the retina:
  - rods and cones
  - light quanta are converted into chemical energy, giving rise to an impulse that is transferred via neurons to the optic nerve.
- Each cone is connected to the optic nerve via a single neuron:
  - spatial acuity of cones is higher than that for rods.
- Several rods are merged into one neuron that connects to the optic nerve:
  - light sensitivity is greater for rods than for cones.
Day and night vision

- The highest spatial acuity of the human visual field is in the **fovea**, a small region of the retina with a visual field of 2 degrees, where the density of cones is largest.

- In **scotopic** (night) vision, luminance levels are low and only rods respond to light:
  - rods are not colour sensitive; therefore objects appear grey at night.

- In **photopic** (day) vision, both rods and cones are activated:
  - photopic vision occurs at luminance levels between 1 and $10^6$ cd/m$^2$. 
18.2 THE HUMAN VISUAL SYSTEM

18.2.2 THE BARTEN MODEL
The human visual system

- 18.2.1 The human eye
- 18.2.2 The Barten model
- 18.2.3 Perceptual linearization
- 18.2.4 Viewing conditions
Barten model

- A model for the contrast sensitivity of the human visual system has been developed by Barten.
- Contrast sensitivity ($S$) is inversely proportional to the threshold modulation ($m_t$).
Threshold modulation

- $m_t$ is the ratio of the minimum luminance amplitude to the mean luminance of a sinusoidal grating such that the grating has a 50% probability to be detected.
Barten model and threshold modulation

- Contrast sensitivity of the human visual system can be measured by presenting images of a sinusoidal grating to a human observer.
- From repeated trials, the threshold modulation can be determined.
- Barten’s model is based on the observation that in order to be detected, the threshold modulation of a grating needs to be larger than that of the internal noise $m_n$ by a factor $k$

\[ m_t = km_n \]
Barten model and MTF

- As the image enters the eye, it is distorted and blurred by the pupil and optical lens.
- This is described by the optical MTF ($M_{opt}$) and is a function of spatial frequency ($u$) and incident luminance ($L$):

$$M_{opt}(u, L) = e^{-2(\pi \sigma(L) u)^2}$$

- The width of $M_{opt}$ depends on pupil diameter which controls the impact of lens aberrations ($C_{ab}$), $\sigma(L)^2 = \sigma_0^2 + (C_{ab} d(L))^2$

- The dependence of pupil diameter ($d$, in mm) on luminance ($L$, in cd/m²) can be approximated by $d(L) = 5 - 3 \tanh(0.4 \log L)$.
Barten model and photon noise

- When light quanta are detected by the retinal cells, photon noise \( F_{ph} \) is incurred with a spectral density given by

\[
\Phi_{ph}(L) = \frac{1}{\eta p E(L)}
\]

where

- \( \eta \) is the quantum detection efficiency of the eye
- \( p \) is the luminous flux to photon conversion factor
- \( E \) is the retinal illumination

\[
E(L) = \frac{\pi d(L)^2}{4} \left( 1 - \left( \frac{d(L)}{9.7} \right)^2 + \left( \frac{d(L)}{12.4} \right)^4 \right)
\]

- \( E(L) \) includes the Stiles-Crawford effect that accounts for variations in efficiency as light rays enter the pupil at different locations
Barten model – lateral inhibition and neural noise

- Lateral inhibition occurs in retinal cells surrounding an excited cell, and is described as

\[ M_{\text{lat}}^2(u) = 1 - e^{-2(u/u_0)^2} \]

- Neural noise (\( F_0 \)) further degrades the retinal signals

- These factors are included in threshold and noise modulation, to give the Barten model modulation
Barten model modulation

\[ m_t M_{opt}(u, L) M_{lat}(u) = 2k \sqrt{\left( \Phi_{ph}(L) + \Phi_{ext} \right) M_{lat}^2(u) + \Phi_0} \]

where

- \( X, Y, T \) are the spatial and temporal dimensions of the object
Complete Barten model for contrast sensitivity of the human eye

The complete Barten model is given by

\[
S(u, L) = \frac{1}{m_t(u, L)} = \frac{M_{opt}(u, L)/k}{\sqrt{\frac{2}{T} \left( \frac{1}{X_0^2} + \frac{1}{X_{\text{max}}^2} + \frac{u^2}{N_{\text{max}}^2} \right) \left( \frac{1}{\eta pE(L)} + \frac{\Phi_0}{M_{\text{lat}}^2(u)} + \Phi_{\text{ext}} \right)}}
\]

where

- \( X_0 \) angular extent of the object
- \( X_{\text{max}} \) maximum integration angle of the eye
- \( N_{\text{max}} \) maximum number of cycles over which the eye can integrate
18.2 THE HUMAN VISUAL SYSTEM

18.2.2 The Barten Model

Contrast sensitivity as a function of luminance

Parameters used were $k=3$, $s_0=8.33\times10^{-3}$ deg, $C_{ab}=1.33\times10^{-3}$ deg/mm, $T=0.1$ s, $X_{\text{max}}=12$ deg, $N_{\text{max}}=15$ cycles, $h=0.03$, $F_0=3\times10^{-8}$ sec/deg$^2$, $u_0=7$ cycles/deg, $X_0=2$ deg, $p=1\times10^6$ phot/sec/deg$^2$/Td

The values for $X_0$ and $p$ correspond to the standard target used in DICOM14
18.2 THE HUMAN VISUAL SYSTEM

18.2.2 The Barten Model

Contrast sensitivity as a function of spatial frequency

Parameters used were $k=3$, $s_0=8.33 \times 10^{-3}$ deg, $C_{ab}=1.33 \times 10^{-3}$ deg/mm, $T=0.1$ s, $X_{\text{max}}=12$ deg, $N_{\text{max}}=15$ cycles, $h=0.03$, $F_0=3 \times 10^{-8}$ sec/deg$^2$, $u_0=7$ cycles/deg, $X_0=2$ deg, $p=1 \times 10^6$ phot/sec/deg$^2$/Td

The values for $X_0$ and $p$ correspond to the standard target used in DICOM14
Experimental verification

- The Barten model as presented here
  - is valid for foveal vision in photopic conditions
  - has been verified with a number of experiments covering the luminance range from 0.0001 to 1000 cd/m²

- Of the parameters in the equation for $S(u,L)$
  - $s_0$, $h$, $k$ were varied to fit measurements
    - $s_0$ varied between 0.45 and 1.1
    - $h$ from 0.005 to 0.3
    - $k$ from 2.7 to 5.0
  - all others were kept fixed
Extensions to Barten model

- Extensions of the model to parafoveal and temporal contrast sensitivity have also been described (Barten 1999)
18.2 THE HUMAN VISUAL SYSTEM

18.2.3 PERCEPTUAL LINEARIZATION
The human visual system

- 18.2.1 The human eye
- 18.2.2 The Barten model
- 18.2.3 Perceptual linearization
- 18.2.4 Viewing conditions
18.2 THE HUMAN VISUAL SYSTEM
18.2.3 Perceptual Linearization

Just Noticeable Difference

Based on the contrast sensitivity $S$, a just noticeable luminance difference (JND) can be defined as twice the luminance amplitude ($a_t$) of a sine-wave grating at threshold modulation.

$$JND(u, L) = 2a_t$$
$$= 2m_t(u, L)L$$
$$= \frac{2L}{S(u, L)}$$
18.2 THE HUMAN VISUAL SYSTEM
18.2.3 Perceptual Linearization

Characteristic curve

- In a display device such as an LCD monitor, the relationship between input gray value $n_g$ and output luminance is non-linear.

- The plot of $n_g$ versus luminance is the so-called characteristic curve of the display device.
Grayscale Standard Display Function (GSDF)

- In a perceptually linearized display, a constant difference of input gray values $\Delta n_g$, results in a luminance difference corresponding to a fixed number $j$ of JND indices, across the entire luminance range of the display

$$\Delta L = j \times JND$$

- To standardize medical devices, medical displays are required to conform to the Grayscale Standard Display Function (GSDF) (DICOM14)
- The GSDF defines the relationship between JND and display luminance for the standard target
  - a sinusoidal pattern of 4 cycles/deg over a 2 deg x 2 deg area
  - embedded in a background with mean luminance equal to the mean luminance of the pattern
Grayscale Standard Display Function (GSDF)
18.2 THE HUMAN VISUAL SYSTEM

18.2.4 VIEWING CONDITIONS
The human visual system

- 18.2.1 The human eye
- 18.2.2 The Barten model
- 18.2.3 Perceptual linearization
- 18.2.4 Viewing conditions
Viewing conditions

- Ambient conditions in a reading room can significantly impact observer performance.
- Ambient lighting decreases performance and is recommended to be kept below 50 lux in a reading room.
- It is also recommended that observers allow time for dark adaptation of their eyes prior to any reading, which takes about 5 minutes.
- Further, there are indications that fatigue deteriorates performance.
18.3 SPECIFICATIONS OF OBSERVER PERFORMANCE

18.3
Specifications of observer performance

- 18.3.1 Decision outcomes
- 18.3.2 Statistical decision theory and ROC methodology
- 18.3.3 Signal-to-noise ratio
18.3 SPECIFICATIONS OF OBSERVER PERFORMANCE

18.3.1 DECISION OUTCOMES
Specifications of observer performance

- 18.3.1 Decision outcomes
- 18.3.2 Statistical decision theory and ROC methodology
- 18.3.3 Signal-to-noise ratio
Binary decision

- A binary decision is exemplified by the task of detecting a signal or abnormality.
- A detection task is a binary decision because there are two truth states:
  - the “normal” state or the absence of a signal ($H_0$)
  - the presence of an abnormality or a signal ($H_1$)
- $H_0$ is often called a negative or disease-free finding.
Possible detection decision outcomes

- Presented with an image $g$ of unknown truth state, the observer must decide whether to categorize $g$ as
  - normal ($D_0$) or
  - as containing a signal ($D_1$)

- There are four possible outcomes of this decision
  - for an image that actually contains a signal
    - a true positive (TP) occurs when it is assigned to $H_1$
    - a false negative (FN) occurs when it is assigned to $H_0$
  - for an image that does not contain a signal
    - a true negative (TN) occurs when it is assigned to $H_0$
    - a false positive (FP) occurs when it is assigned to $H_1$
### 18.3 SPECIFICATIONS OF OBSERVER PERFORMANCE

#### 18.3.1 Decision outcomes

#### Possible detection decision outcomes

<table>
<thead>
<tr>
<th>Actual condition</th>
<th>$H_0$: signal absent (negative)</th>
<th>$H_1$: signal present (positive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- D₀: signal absent (negative)  
  - TN  
  - FN Type II error

- D₁: signal present (positive)  
  - FP Type I error  
  - TP

**Possible detection decision outcomes**

- **Decision**: H₀: signal absent (negative)  
  - TN  
  - FN Type II error

- **Decision**: H₁: signal present (positive)  
  - FP Type I error  
  - TP

**Diagnostic Radiology Physics: A Handbook for Teachers and Students – 18.3.1 Slide 5 (41/110)**
Decision errors

- The observer can make two types of decision errors
- A **type I error** occurs when the truth state $H_0$ is rejected when it is actually true (FP)
- A **type II error** occurs when the truth state $H_0$ is not rejected when it is not true (FN)

In medical applications, the cost associated with a

- type I error results in patient anxiety and societal costs because of additional tests and procedures
- type II error is a missed cancer or misdiagnosis
18.3 SPECIFICATIONS OF OBSERVER PERFORMANCE
18.3.1 Decision outcomes

Accuracy

- The accuracy of a medical procedure is given by the correct decisions per total number of cases ($N$)

\[
\text{accuracy} = \frac{TP + TN}{N}
\]

- Accuracy does not account for prevalence and can therefore be misleading
  - in screening mammography, the cancer prevalence in some screening populations is about 4 per 1000 patients
  - the accuracy of a radiologist who classifies all screening cases as normal is 9996/10000 or 99.96%
  - such a radiologist would have missed all cancers!

- More meaningful performance measures are sensitivity, specificity and associated variables
Sensitivity

- **Sensitivity**, also known as the true positive fraction (TPF), measures the proportion of actual true (positive) cases that are correctly identified.

\[
TPF = \frac{TP}{(TP + FN)}
\]
Specificity

- Specificity measures the proportion of negative cases that are correctly identified.
- It is conveniently derived through the false positive fraction (FPF) where

\[
\text{specificity} = 1 - \text{FPF}
\]

and

\[
\text{FPF} = \frac{FP}{TN + FP}
\]
PPV and NPV

- Often in the literature, positive predictive values (PPV) and negative predictive values (NPV) are also reported.

\[
PPV = \frac{TP}{TP + FP}
\]

\[
NPV = \frac{TN}{TN + FN}
\]
18.3 SPECIFICATIONS OF OBSERVER PERFORMANCE

18.3.2 STATISTICAL DECISION THEORY AND ROC METHODOLOGY
Specifications of observer performance

- 18.3.1 Decision outcomes
- 18.3.2 Statistical decision theory and ROC methodology
- 18.3.3 Signal-to-noise ratio
Decision theory

- In a probabilistic approach to perception, which is also referred to as decision theory, the observer derives a decision variable $\lambda$ from each image.

- The conditional probability density functions (PDFs) of the decision variable $\lambda$ under the truth states $H_0$ and $H_1$, $p(\lambda|H_0)$ and $p(\lambda|H_1)$, are shown on the next slide.
Decision outcomes in the probabilistic decision paradigm
18.3 SPECIFICATIONS OF OBSERVER PERFORMANCE

18.3.2 Statistical decision theory and ROC methodology

Decision outcomes in the probabilistic decision paradigm

- The operating point of the observer, $\lambda_c$, is the decision threshold based on which an observer will call an image “normal” (negative) or “abnormal” (positive)
- Both the TPF and FPF values depend on $\lambda_c$
18.3 SPECIFICATIONS OF OBSERVER PERFORMANCE
18.3.2 Statistical decision theory and ROC methodology

Binormal model

- Often it is assumed that both PDFs are normally distributed so that a binormal model can be used

- If
  - $p(\lambda|H_0)$ has zero mean and unit variance
  - $p(\lambda|H_1)$ has a mean of $a/b$ and variance $1/b$

- The FPF and TPF are given by

$$FPF(\lambda) = \Phi(-\lambda) \quad TPF(\lambda) = \Phi(a - b\lambda)$$

where

- $\Phi(.)$ is the cumulative normal distribution function $\Phi(y) = \int_{-\infty}^{y} \Phi(x)dx$
- $\phi(.)$ is the normal PDF $\phi(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$
Location of decision thresholds in binormal model

- Shows the probability density functions associated with $H_0$ and $H_1$
- Different $\lambda_c$ describe the vigilance of the reader
18.3 SPECIFICATIONS OF OBSERVER PERFORMANCE

18.3.2 Statistical decision theory and ROC methodology

Vigilance of the reader

- The choice of operating threshold depends on
  - the task
  - the costs associated with decision errors

- A low decision threshold ($\lambda_c = A$) corresponds to an “aggressive” reader
  - characterised by a high sensitivity but a low specificity
  - appropriate to a diagnostic task

- Higher decision thresholds (operating points B or C)
  - when screening for a condition with low prevalence
  - the reader typically operates at a high specificity (FPF < 0.1) in order to reduce the number of callbacks to an acceptable level
  - the associated decrease in sensitivity might be offset by the yearly repeat of the screening procedure
Receiver-operating characteristic (ROC) curve

- The receiver-operating characteristic (ROC) curve reveals the trade-off between sensitivity and specificity.
- This very useful and much-used curve is generated by plotting the TPF versus the FPF for all values of the decision threshold $\lambda$. 
18.3 SPECIFICATIONS OF OBSERVER PERFORMANCE

18.3.2 Statistical decision theory and ROC methodology

Receiver-operating characteristic (ROC) curve

- The end points of the ROC curve are (0,0) and (1,1)
- The ROC curve lies on or above the diagonal
- If an ROC curve is below the diagonal it implies that the truth states $H_0$ and $H_1$ have been interchanged, which can be caused by an observer misreading instructions
- If the PDFs have equal variance under both truth states, the ROC curve is symmetrical about a line from (0,1) to (1,0)

- ROC curves for $a = 2, 1.3, 0.4, 0$; $b=1$ in binormal model
- The underlying PDFs have equal variance and therefore the ROC curves are symmetrical
- The operating points A, B and C correspond to the decision thresholds shown on earlier slide
Area under ROC curve (AUC)

- The area under the ROC curve, AUC, quantifies overall decision performance and is the integral of TPF over FPF

\[ AUC = \int_{0}^{1} TPF(FPF) dFPF \]

- The values of AUC vary between 1.0 and 0.5
- A perfect decision maker achieves an AUC of 1.0, while random guessing results in AUC of 0.5
- Within the binormal model, AUC is given by

\[ AUC = \Phi \left( \frac{a}{\sqrt{1 + b^2}} \right) \]
Comparing two procedures

- AUC is an overall measure of decision performance that does not provide any information about specific regions of the ROC curve
  - which may be important when deciding between two procedures that are to be used in a specific task

- For example, consider two procedures are to be evaluated for their performance in a screening task
Example: Comparing two procedures for a screening task

- The AUC for both procedures is 0.82
- However, in the region of specificity < 0.9 the performance of procedure A exceeds that of procedure B
- Specificity < 0.9 typically would be a limit when screening for a condition with low prevalence such as breast cancer

- ROC curves for two procedures with equal AUC, but different $a$ and $b$
18.3 SPECIFICATIONS OF OBSERVER PERFORMANCE

18.3.3 SIGNAL-TO-NOISE RATIO
Specifications of observer performance

- 18.3.1 Decision outcomes
- 18.3.2 Statistical decision theory and ROC methodology
- 18.3.3 Signal-to-noise ratio
A frequently used measure of performance is the signal-to-noise ratio (SNR), estimated from the decision variables $\lambda_0$ and $\lambda_1$ under the truth states $H_0$ and $H_1$ through

$$SNR = \frac{\langle \lambda_1 \rangle - \langle \lambda_0 \rangle}{\sqrt{1/2(\text{var}(\lambda_0) + \text{var}(\lambda_1))}}$$

This expression is valid only if
- the decision variables are normally distributed
- the distributions of $\lambda$ are fully characterized by mean and variance

If the distributions depart from normality, SNR values computed by this expression can be misleading.
SNR and AUC

- SNR is related to AUC through

\[ AUC = \frac{1}{2} \left(1 + Erf \left(\frac{SNR}{2}\right)\right) \]

where \( Erf \) is the Gaussian error function

\[ Erf(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \]

- For a detection task, SNR is often called detectability and labelled \( d' \)

- \( d' \) may be estimated directly from the image statistics, measured experimentally
18.4 EXPERIMENTAL METHODOLOGIES
18.4 EXPERIMENTAL METHODOLOGIES

18.4 Experimental methodologies

- 18.4.1 Contrast-detail methodology
- 18.4.2 Forced choice experiments
- 18.4.3 ROC experiments
Quantifying image quality

- The preferred metric for quantifying image quality of a medical imaging system is **task performance**
  - measures the performance of an observer in a well-defined task based on a set of images

- Image quality is thus a statistical concept

- Three methodologies to measure performance of a human reader (observer) in the task of detecting a known signal in a noisy background:
  - 18.4.1 Contrast-detail methodology
  - 18.4.2 Forced choice experiments
  - 18.4.3 ROC experiments
18.4 EXPERIMENTAL METHODOLOGIES

18.4.1 CONTRAST-DETAIL METHODOLOGY
Experimental methodologies

- 18.4.1 Contrast-detail methodology
- 18.4.2 Forced choice experiments
- 18.4.3 ROC experiments
Contrast-detail methodology

- To provide an objective measure of image quality, imaging systems have been evaluated using **contrast-detail phantoms**.
- Consist of disks of increasing diameter and contrast arranged along columns and rows.
- The reader’s task is to identify the lowest contrast where the signal can be perceived, for each detail size.
- The reader’s threshold contrasts, plotted as a function of disk radius, is called a **contrast-detail curve**.
18.4 EXPERIMENTAL METHODOLOGIES
18.4.1 Contrast-detail methodology

Contrast-detail phantom

- The CDMAM phantom is an example of a contrast-detail phantom
- Consists of gold disks of exponentially increasing diameter (0.06 to 2 mm) and thicknesses (0.002 to 0.03 mm)
**Drawbacks of contrast-detail methodology**

- Threshold contrast is dependent on the internal operating point of each observer
  - inter- and intra-observer variabilities can be quite high
- There may be a memory effect where the observer anticipates and therefore reports a signal, while the signal cannot yet be perceived
Overcoming drawbacks of contrast-detail methodology

- The CDMAM phantom has been developed to address some of these shortcomings (Bijkerk et al 2000)
- Places two signals
  - one at the centre
  - a second randomly in one of the four corners of each cell in the array
- This arrangement prevents the reader from anticipating the signal
- An alternative approach is to determine contrast thresholds using computer reading of the images
18.4 EXPERIMENTAL METHODOLOGIES
18.4.2 FORCED CHOICE EXPERIMENTS
18.4 EXPERIMENTAL METHODOLOGIES

18.4

Experimental methodologies

- 18.4.1 Contrast-detail methodology
- 18.4.2 Forced choice experiments
- 18.4.3 ROC experiments
Alternative-forced-choice (AFC) experiments

- A different way of measuring human observer performance is through alternative-forced-choice (AFC) experiments.
- In an AFC experiment, $M$ alternative images are shown to the observer, whose task is to identify the image that contains the signal.
- The proportion of correct answers ($PC$) in an $M$-AFC experiment is computed through

$$PC = \frac{\text{number of correctly scored trials}}{\text{total number of trials}}$$
18.4 EXPERIMENTAL METHODOLOGIES
18.4.2 Forced choice experiments

Number of trials required

- The value of $PC$ varies between 1 and the guessing score $1/M$
- The number of trials (that is, the number of test images shown) determines the accuracy of $PC$
- Typically, more than 100 trials are required
**18.4 EXPERIMENTAL METHODOLOGIES**

**18.4.2 Forced choice experiments**

*PC* in terms of PDF

- *PC* can be predicted based on the assumption that the observer will assign the score to the image which evokes the highest internal response
- If the probability density function of negative responses is normal with zero mean and unit variance
- Then the probability that the response of the *M*-1 actually negative responses exceed that of the actually positive response (normally distributed with mean *d'* and unit variance) becomes

\[
PC = \int_{-\infty}^{\infty} \Phi(x)^{M-1} \phi(x - d') dx
\]

- where \(\phi(.)\) and \(\Phi(.)\) are the normal PDF and cumulative normal distribution function as previously defined

*Eckstein & Whiting, 1996*
18.4 EXPERIMENTAL METHODOLOGIES

18.4.2 Forced choice experiments

**PC and AUC**

- Using

\[ FPF(\lambda) = \Phi(-\lambda) \]

\[ TPF(\lambda) = \Phi(a - b\lambda) \]

\[ AUC = \int_{0}^{1} TPF(FPF)dFPF \]

- And

\[ PC = \int_{-\infty}^{\infty} \Phi(x)^{M-1} \phi(x - d')dx \]

- It can be shown that when \( M=2 \)

\[ PC = AUC \]
2-AFC or $M$-AFC experiments

- Typically 2-AFC experiments are carried out so that the observers score 90% of all trials correctly.
- Depending on the signal, the corresponding amplitude thresholds may need to be very low, which can strain the observer.
- By increasing the number of alternative locations ($M$), the task becomes more difficult and larger threshold amplitudes are required.
- Experiments with $M$ between 2 and 1800 have been conducted (Burgess & Ghandeharian 1984).
**M-AFC experiments**

- The alternative signal locations
  - can be centres of individual ROIs
  - or all possible signal locations can be contained within one single test image

- When a single test image is used
  - must ensure that the distance between locations is larger than the correlation distance of the background
  - otherwise observer responses might become correlated

- If the responses are uncorrelated, $d'$ can be determined from $PC$ by inverting

$$PC = \int_{-\infty}^{\infty} \Phi(x)^{M-1} \phi(x - d') dx$$
18.4 EXPERIMENTAL METHODOLOGIES
18.4.3 ROC EXPERIMENTS
18.4 EXPERIMENTAL METHODOLOGIES

18.4

Experimental methodologies

- 18.4.1 Contrast-detail methodology
- 18.4.2 Forced choice experiments
- 18.4.3 ROC experiments
In an ROC experiment, a single image is presented to the observer whose task is to provide a likelihood rating. In a detection experiment, this could be the “likelihood of signal present.” Researchers have used:

- continuous rating scales
- categorical scales
  - typically fewer than 10 categories are provided
18.4 EXPERIMENTAL METHODOLOGIES
18.4.3 ROC experiments

ROC vs. 2-AFC experiments

- For a given number of images available, ROC experiments provide better statistical power because
  - one single image is required per trial
  - in a 2-AFC experiment, two images are required per trial

- Furthermore, a ROC curve can be generated from rating data, which provides more detailed information on task performance in specific regions

- On the other hand, ROC experiments are more demanding of the observer, which can increase reading time and observer fatigue
ROC curve fitting

- Curve fitting problems, such as the presence of hooks, can be overcome by using ROC curve fitting software provided by the University of Chicago
  - [http://www-radiology.uchicago.edu/krl/KRL_ROC/software_index6.htm](http://www-radiology.uchicago.edu/krl/KRL_ROC/software_index6.htm)

- Their curve fitting model does not allow for the ROC curve to fall below the diagonal.
Search experiments

- A drawback of ROC/2-AFC experiments is the lack of signal localization.

- In a clinically realistic task such as screening for cancer, the radiologist is required to indicate locations of potential lesions, in addition to his overall rating of the image.

- To allow for more clinically realistic laboratory experiments, several extensions to ROC have been developed:
  - LROC
  - FROC/AFROC/JAFROC
Search experiments – LROC

- In the **Location ROC (LROC)** methodology, the observer is required:
  - to indicate the location of a lesion and
  - to provide a score

- In an LROC curve, the false positive fraction is plotted along the x-axis, but on the y-axis, the true positive fraction with correct signal location is plotted.

- The upper right endpoint of the LROC curve is determined by the proportion of correctly located signals.
Search experiments – FROC

- In the Free-Response ROC (FROC) experiments, the observer
  - indicates the location of a lesion, but
  - provides no rating

- The FROC curve is the proportion of correctly detected signals plotted as a function of the average number of false positive detections per image

- FROC analysis is often used in the performance assessment of computer-aided detection schemes
Search experiments – AFROC and JAFROC

- In Alternative Free-Response ROC (AFROC) methodology, the proportion of correctly detected signals is plotted as a function of the probability that at least one false positive per image is found.

- In the AFROC methodology, a summary figure-of-merit exists, $A_{IJ}$.

- This is the probability that lesions are rated higher than false positive marks in normal images.

- JAFROC is a software package developed to estimate $A_{IJ}$ and statistical significance from human observer FROC data (Chakraborty website).
18.5 OBSERVER MODELS
18.5 OBSERVER MODELS

- Often it is impractical to determine task performance of imaging systems from studies involving human observers, simply because reader time is scarce and expensive, especially for a trained reader such as a radiologist.

- To address this issue, mathematical observer models have been developed, either to predict human performance or to estimate ideal observer performance.

- The observer models described here assume that images contain exactly known additive signals:
  - i.e. signals of known shape and amplitude that have been added to a noisy background.
18.5 OBSERVER MODELS

18.5.1 THE BAYESIAN IDEAL OBSERVER
The Bayesian ideal observer

- In a detection task, the observer is faced with the decision of assigning the unknown data $g$ to one of the truth states.
- The Ideal Bayesian Observer’s decision strategy is to minimize the average cost $<C>$ of the decision.
- With the observer’s decision outcomes $D_0$ and $D_1$, the average cost is

$$<C> = C_{00} P(D_0 | H_0) P(H_0) + C_{11} P(D_1 | H_1) P(H_1)$$

$$+ C_{01} P(D_0 | H_1) P(H_1) + C_{10} P(D_1 | H_0) P(H_0)$$

where

- $C_{00}$ and $C_{11}$ are costs associated with correct decisions.
- $C_{01}$ is the cost of a false negative decision.
- $C_{10}$ is the cost of a false positive decision.
18.4 OBSERVER MODELS

18.5.1 The Bayesian ideal observer

The Bayesian ideal observer's decision rule

Minimizing average cost leads to the decision rule

\[
\text{choose } H_1 \text{ if } \Lambda(g) = \frac{p(g | H_1)}{p(g | H_0)} \geq \frac{P(H_0)(C_{10} - C_{00})}{[1 - P(H_0)](C_{01} - C_{11})} \\
\text{else} \\
\text{choose } H_0
\]
The Bayesian criterion

- This decision criterion is the so-called Bayes criterion
- The resulting minimum cost is the Bayes risk
- A test based on the ratio of probabilities is called a likelihood ratio test
- It can be shown that the likelihood ratio test maximizes sensitivity for a given false-positive fraction
- Therefore the ideal observer maximizes the AUC and provides an upper limit on task performance
Log-likelihood ratio

- Knowledge of the probability density functions $p(g|H_i)$ is required to form the decision variable (also called test statistic) $\Lambda(g)$

- $\Lambda(g)$ is generally difficult to determine for realistic imaging tasks

- When the underlying data statistics are normal, it is often easier to work with the log-likelihood ratio, $\lambda(g) = \log(\Lambda(g))$

- Since the logarithmic transformation is monotonic, it does not affect the decision outcome or AUC
18.5 OBSERVER MODELS

18.5.2 OBSERVER PERFORMANCE IN UNCORRELATED GAUSSIAN NOISE
18.4 OBSERVER MODELS
18.5.2 Observer performance in uncorrelated Gaussian noise

Observer performance in uncorrelated Gaussian noise

- If the image noise is
  - zero-mean Gaussian distributed with variance $\sigma^2$
  - independent in each image pixel

- The ideal observer derives the decision variable $\lambda$ by convolving each image $g$ with a template $w$

$$\lambda = w^t \cdot g$$

- The ideal observer template for detection of additive, exactly known signals is the signal itself, $w = s$
18.5.2 Observer performance in uncorrelated Gaussian noise

Estimation of $d'$ (SNR / detectibility)

- From the test statistic $\lambda$, the SNR $d'$ can be estimated using

$$SNR = \frac{<\lambda_1> - <\lambda_0>}{\sqrt{1/2(\text{var}(\lambda_0) + \text{var}(\lambda_1))}}$$

- If the image statistics are known, $d'$ can be computed directly using

$$d'^2 = \frac{SE}{\sigma^2}$$

where

- the signal energy for a signal containing $N$ pixels is $SE = \sum_{i=1}^{N} s_i^2$
Observer model for human performance in uncorrelated Gaussian noise

- Compared to the ideal observer, a human reader performs sub-optimally.
- To account for the signal degradation by the human visual system, the template $w$ can be formed by convolving the signal with an eye filter $e$

$$w = e \ast s$$

where

- $e$ is defined in the spatial frequency domain as

$$E(f) = fe^{-cf}$$
18.4 OBSERVER MODELS
18.5.2 Observer performance in uncorrelated Gaussian noise

Non-prewhitening observer with eye filter (NPWE)

- Typically, $c$ is chosen so that $E(f)$ peaks at 4 cycles/mm
- This model observer model is called a non-prewhitening observer with eye filter (NPWE)
- This observer model
  - can predict the performance of a human reader when the image background is uniform (Segui & Zhao 2006)
  - fails when strong background correlations are present, such as in anatomic backgrounds (Burgess 1999, Burgess et al 2001)
18.5 OBSERVER MODELS

18.5.3 OBSERVER PERFORMANCE IN CORRELATED GAUSSIAN NOISE
Observer performance in correlated Gaussian noise

- When the noise in the image is Gaussian correlated, the background statistics can be characterized by the covariance matrix $K_g$

\[
K_g = \langle (g - \langle g \rangle)(g - \langle g \rangle)^t \rangle
\]

where

- $\langle . \rangle$ indicates the expectation value

- In this background type, the ideal observer pre-whitens the image by using a template that removes the background correlation

\[
w = K_g^{-1}s
\]
18.4 OBSERVER MODELS
18.5.3 Observer performance in correlated Gaussian noise

Estimation of $d'$ (SNR / detectability)

- For ideal observer
  
  $$d'^2 = s^t K_g^{-1} s$$

- Estimating the background covariance $K_g$ is often difficult
  
  • but when the image data are stationary, the covariance matrix is diagonalized by the Fourier transform
  
  • allows computation of $d'$ using

  $$d'^2 = \int \frac{|S(f)|^2}{W(f)} df$$

  where
  
  • $S(f)$ is the Fourier transform of the signal $s$
  
  • $W(f)$ is the background power spectrum

- This equation is the basis of estimation of SNR from noise-equivalent quanta in chapter 4 (image quality)
Channel functions

- Another approach to estimating the background covariance is to expand the image into so-called channels or basis functions.

- The choice of channel functions depends on the objective of the observer model.

- It has been found that:
  - Laguerre-Gauss are efficient channel functions when approximating the ideal observer.
  - whereas Gabor channels provide a good estimate of human observer performance.
18.4 OBSERVER MODELS

18.5.3 Observer performance in correlated Gaussian noise

Channelized Hotelling observers

- Often, 10 channel functions or less are sufficient to represent $g$
  - they reduce the dimensionality of the problem considerably, without the assumption of background stationarity
- These observer models are often called **channelized Hotelling observers**
- The University of Arizona provides software and tutorials for assessing image quality by use of observer models
  - [http://www.radiology.arizona.edu/CGRI/IQ/index.html](http://www.radiology.arizona.edu/CGRI/IQ/index.html)
18. BIBLIOGRAPHY
18. BIBLIOGRAPHY

18. BIBLIOGRAPHY

- CHAKRABORTY, D., FROC web site


18. BIBLIOGRAPHY

INTERNATIONAL COMMISSION ON RADIATION UNITS AND MEASUREMENTS, Receiver Operating Characteristic Analysis in Medical Imaging, ICRU Rep. 79 Bethesda, MD (2008)

SEGUI J. and ZHAO W., Amorphous selenium flat panel detectors for digital mammography: Validation of a NPWE model observer with CDMAM observer performance experiments, Med. Phys. 33 (2006), 3711-3722