9.63 Laboratory in Visual Cognition

Fall 2009

Mon. Sept 14

Signal Detection Theory

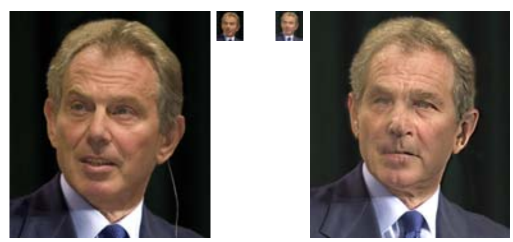
Map of the course

- I - Signal Detection Theory : Why it is important, interesting and fun?
- II - Some Basic reminders
- III - Signal Detection Theory: Core
- IV – Signal Detection Theory: Advanced

Who is that person?

Close-up, George Bush and from far away, Tony Blair.

Signal Detection: Merging Signals



Depending on the distance of viewing (or the size of the image), you will see one face or the other. Here the distribution of "face features" of Bush and Blair overlap: it is difficult to extract the signal corresponding to the face of Blair from the other signal, the face of Bush.

Detecting Emotion and Attitude

Which face is *positive* vs. *negative*?

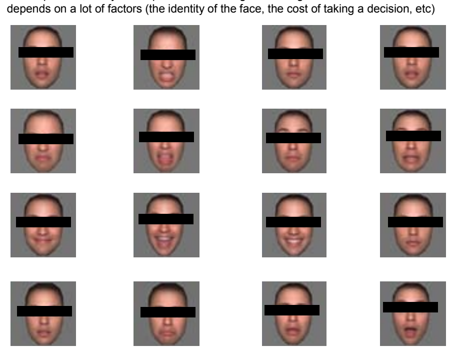
fear

surprised

Figures by MIT OpenCourseWare.

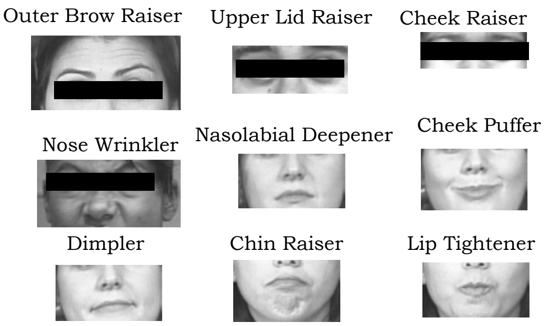
Positive or Negative Emotion?

For each face, there are 4 outcomes: the face is negative or the face is positive, and your response could be right or wrong. Your decision depends on a lot of factors (the identity of the face, the cost of taking a decision, etc)



Figures by MIT OpenCourseWare.

Face Parts Vocabulary

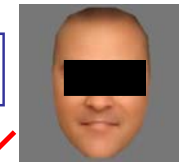


Figures by MIT OpenCourseWare.

Tian, Kanade, Cohn (2001). Recognizing Action units for Facial Expression Analysis. IEEE PAMI, 23,2,1-19.

Face understanding is about making the right decision

After hearing the news (one is promoted to Singapore and one is fired) is that person upset or not ?



Figures by MIT OpenCourseWare.

	The person is truly "angry", "upset"	The person is not upset
You think the person is upset	Hit	False Alarm (false positive)
You think the person is happy	Miss (false negative)	Correct Rejection

Categorical boundary

Positive or negative?

Figures by MIT OpenCourseWare.

Signal Detection Theory is everywhere

Figure removed due to copyright restrictions.

- If a person is trying to detect very weak signals in a background of noise-for example, picking out blips on a radar screen-the problem confronting the person is to *pick out the signal from the noise*.
- But if the signal is very faint, or the noise level is very high, the observer might make errors.
- There are two types of errors a person can make: false positives and false negatives.

false positives (false alarms)

Figure removed due to copyright restrictions.

- 1. *False positives* occur if a person says *yes* (a positive response) but this is wrong (false) because no signal was presented. A *false positive* response can also be called a *false alarm*. If you thought you heard somebody call your name, but nobody actually did, that is a false positive.

Prosopagnosia, amnesia

false negatives

- *False negatives* occur when a person says *no* (a negative response) but this is false because actually a signal *was* presented.
- An example would be failing to detect a blip on a radar screen.

Figure removed due to copyright restrictions.

When false negatives cost a lot..

Figure removed due to copyright restrictions.

- In some situations, false positives (or false alarm) do not cause as much harm as false negatives.
- Consider a blood bank screening samples for the AIDS (HIV) virus. The initial screening of blood samples uses a very sensitive test designed to eliminate false negatives, even though that means there will be some false positives.
- False positives, in this case, are blood samples which test positive for the HIV virus, although later testing shows they are not really infected.
- These false positives have a cost—some blood is wasted—but that is a small price to pay for the security of knowing that no infected blood is given to hospital patients who receive transfusions. In other words, there must be no false negatives in this situation.

When false positives cost a lot..

- In other situations, a more important goal is to *avoid false positives* at all costs.
- A hunter must learn *not* to shoot at everything that moves in the bushes, because the moving object might be a human or a dog. The hunter must wait until the form of the object becomes clear. The threshold for pulling the trigger and shooting must be raised so that a signal does not lead to a response unless the signal is clearly perceived.
- False negatives (failing to shoot) are less important than false positives (shooting the wrong thing).

A tumor scenario

- You are a radiologist examining a CT scan, looking for a tumor.
- The task is hard, so there is some uncertainty: either there is a tumor (signal present), or there is not (signal absent). Either you see a tumor (response “yes”), either you do not see a tumor (response “no”).
- There are 4 possible outcomes:

	Signal Present	**Signal Absent**
Say “Yes"	**Hit**	**False Alarm**
Say “No"	Miss	Correct Rejection

Figures removed due to copyright restrictions.

Adapted from David Heeger document

Decision making process

- Two main components:
- **(1) The signal or information**. You look at the information in the CT scan. A tumor might be brighter or darker, have a different texture, etc.

  With expertise and additional information (other scans), the likelihood of getting a HIT or CORRECT REJECTION increase.

Figure removed due to copyright restrictions.

Adapted from David Heeger document

Decision making process

- Two main components:
- (2) **Criterion**: the second component of the decision process is very different: it refers to **your own judgment** or “internal criterion”. For instance, for two doctors:
- Criterion “life and death” (and money): Increase in False Alarm = decision towards “yes” (tumor present) decision. A false alarm will result in a routine biopsy operation.
- **This doctor has a bias toward “yes”: liberal response strategy**.
- Criterion “unnecessary surgery”: surgeries are very bad (expensive, stress). They will miss more tumors and save money to the social system. They will feel that a tumor if there is really one will be picked-up at the next check-up.
- **This doctor has a bias towards “no”: a conservative response strategy.**

Adapted from David Heeger document

Figure removed due to copyright restrictions.

Is there a tumor ?

Figures removed due to copyright restrictions.

Hypothetical internal response curves

- Starting point of signal detection theory: all reasoning and decision making takes place in the presence of **uncertainty**. For each decision, you have a certain level of "internal" certainty.
- Your internal response depends on the visual information (stimulus intensity) and the state of your sensory system (which varies for a given observer (tired, confident, more or less attention, too much caffeine).
- SDT assumes that the sensory system of the observer is "noisy", meaning it is subject to random fluctuations that are unrelated to the stimulus event. Even if the visual stimulus is the same, the sensory signal will not be the same on every trials.
- SDT assumes that the sensory response will vary randomly over trials around an average value, producing **a normal distribution of strengh values (of internal response).**

Figure removed due to copyright restrictions.

Figure 1: Internal response probability of occurrence curves for noise-alone and for signal-plus-noise trials.

Adapted from David Heeger document

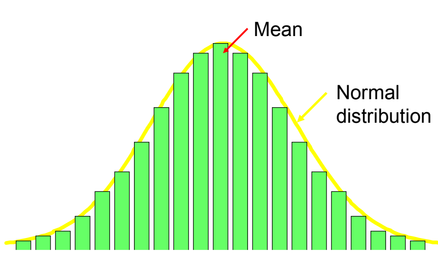
Signal Detection Theory: Definition

- **Signal detection theory**, is a means to quantify the ability to discern between a signal and the absence of signal (or noise)
- Your decision depends on the signal but also your **response bias**

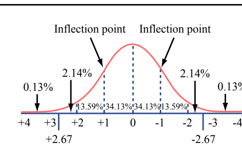
Map of the course

- I - Signal Detection Theory : Why it is important, interesting and fun?
- II - Some Basic reminders
- III - Signal Detection Theory: Core
- IV – Signal Detection Theory: Advanced

Normal Distribution



Normal Distribution



Proportions of scores in specific areas under the normal curve. The inflection points are one standard deviation from the mean.

Figure by MIT OpenCourseWare.

- Each side of the normal curve has a point where the curve slightly reverses its direction: this is the inflection point
- **The inflection point is always one standard-deviation from the mean**
- About **68 %** of all scores are contained within one standard deviation of the mean
- 96 % of the scores are contained within 2 stdev
- 99.74 % of the scores are contained within 3 stdev
- This property of normal curves is extremely useful because if we know an individual's score and the mean and standard deviation in the distribution of scores, we also know the person relative rank.

IQ Population Distribution

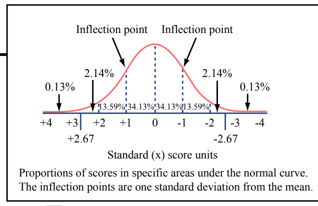
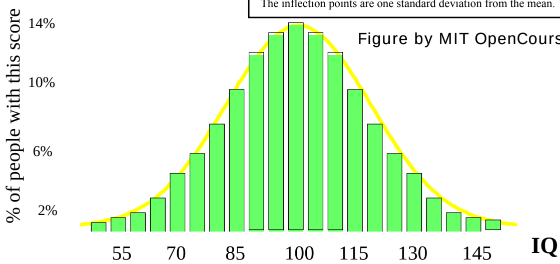


Figure by MIT OpenCourseWare.



Most IQ tests are devised so that the population mean is 100 and the stdev is 15.
If a person has an IQ of 115, she has scored higher than **XX** % of all people.

Normal Curve and Z scores

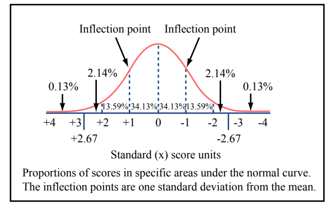


Figure by MIT OpenCourseWare.

- It is common to compare scores across normal distributions with different means and variances in terms of standard scores or **Z-scores**
- The z-score is the difference between an individual score and the mean expressed in units of standard deviations.

Map of the course

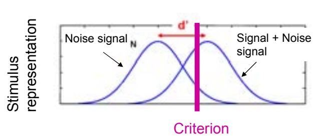
- I - Signal Detection Theory : Why it is important, interesting and fun?
- II - Some Basic reminders
- III - Signal Detection Theory: Core
- IV – Signal Detection Theory: Advanced

Signal Detection: Intuition

- Consider 2 participants in a visual detection task of a faint target. The researcher asks the participants to report if they saw the target.
- After 50 trials, participant A reports seeing the target 25 times
- Participant B reports detection 17 times.
- Did participant A do better than B?

Signal Detection Theory (SDT)

- The underlying model of SDT consists of **two normal distributions**
- One representing a signal (target present)
- The other one representing the "noise" (target absent)
- The willingness of the person to say 'Signal Present' in response to an ambiguous stimulus is represented by the **criterion**.



Internal Response and Internal noise

"The internal response refers to your internal impression or "state of mind".

Figure by MIT OpenCourseWare.

- Let's suppose that our doctor has a set of tumor detector neurons and that we monitor the response of one of these neurons to determine the likelihood that there is a tumor in the image.
- These hypothetical tumor detectors will give noisy and variable responses. After one glance at a scan of a healthy lung, our hypothetical tumor detectors might fire 10 spikes per second.
- After a different glance **at the same scan** and under the same conditions, these neurons might fire 40 spikes per second.
- This internal response is inherently noisy and determine the doctor's impression. The internal response is in some unknown, but in theory, quantifiable, units.

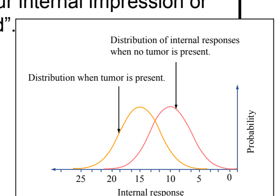


Figure 1: Internal response probability of occurrence curves for noise-alone and for signal-plus-noise trials.

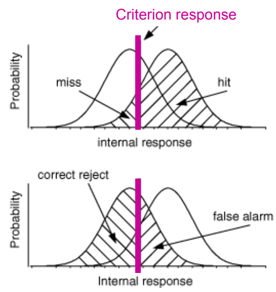
Figure shows a graph of two hypothetical internal response curves. The curve on the left is for the *noise-alone* (healthy) trials, and the curve on the right is for the *signal-plus-noise* (tumor present) trials.

The horizontal axis is labeled *internal response* and the vertical axis is labeled *probability*. The height of each curve represents how often that level of internal response will occur.

Text from David Heeger document

Hypothetical internal response curves

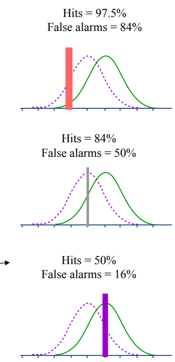
- **SDT assumes that your internal response will vary randomly over trials around an average value, producing a normal curve distribution of internal responses.**
- The decision process compare the strength of the internal (sensory) response to a criterion: whenever the internal response is greater than this criterion, response "yes". Whenever the internal response is less than the criterion, response "no".
- The decision process is influenced by knowledge of the probability of signal events and payoff factors (e.g. motivation)



Adapted from David Heeger document

Effects of Criterion

- If you choose a "low" criterion, you respond "yes" to almost everything (never miss a tumor and have a very high % correct – HIT - rate, but a lot of unnecessary surgeries).
- If you choose a high criterion, you respond "no" to almost everything.



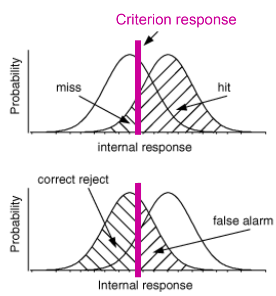
from David Heeger document

Figure by MIT OpenCourseWare.

Criterion and Type of Responses

- Criterion line divides the graph into 4 sections (hits, misses, false alarm, correct rejections).

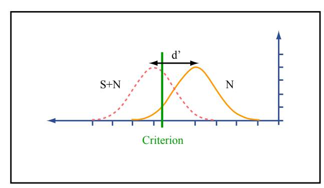
	Signal Present	Signal Absent
Say "Yes"	Hit	False Alarm
Say "No"	Miss	Correct Rejection



Adapted from David Heeger document

SDT and d-prime

- How well a person can discriminate between Signal Present and Signal Absent trials is represented by the difference between the means of the two distributions, d'.



http://wise.cgu.edu/sdt/models_sdt1.html

Figure by MIT OpenCourseWare.

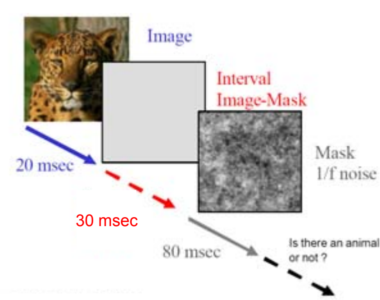
I have data from an experiment with target present/absent: Where do I start to calculate d' ?

- Everything is in the HIT and FA rates. For instance:

HIT = 0.84 **FA = 0.16**

- Response bias c = -0.5[z(H)+z(F)]
- Sensitivity (discrimination) d'= z(H)-z(F)

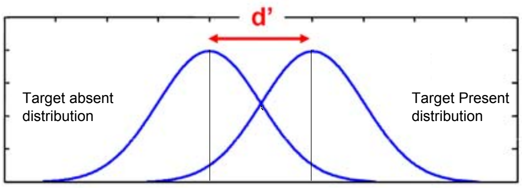
Detection Task - Method



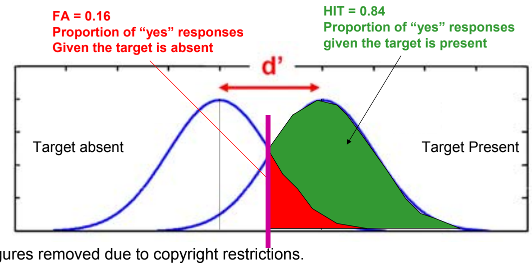
D-prime: d' = separation / spread

$d' = z(\text{hit rate}) - z(\text{false alarm})$

Figures removed due to copyright restrictions.



D-prime: d' = separation / spread

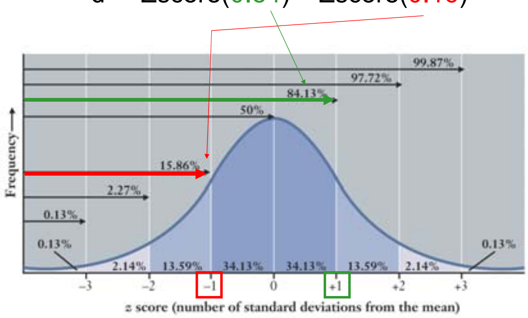


Figures removed due to copyright restrictions.

Criterion

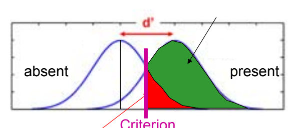
d' = Zscore(HIT) – Zscore(FA)

d' = Zscore(0.84) – Zscore(0.16)



Figures removed due to copyright restrictions.

HIT = 0.84 = Area is 84 %



FA = 0.16
Area is 16 %

For calculation of d', consider the % of data of a normal curve

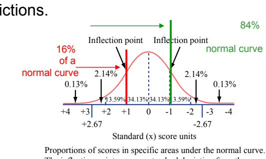


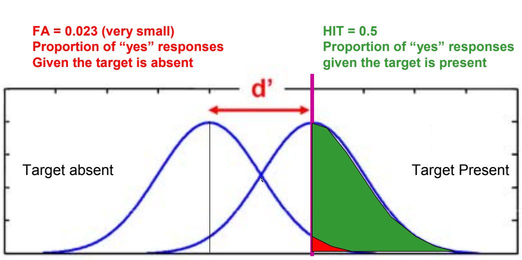
Figure by MIT OpenCourseWare.

Response bias c = -0.5[z(H)+z(F)]
C = - 0.5 [z(0.84) + z(0.16)]
C = 0 (no bias)

Sensitivity d'= z(H)-z(FA)
d'= z(0.84) - z(0.16)
d'= 1 – (-1)
d' = 2

In Excel: z(X) is NORMINV(x,0,1) e.g. NORMINV(0.84,0,1)

D-prime: d' = separation / spread

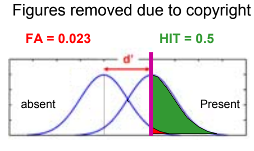


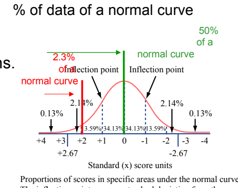
Higher Criterion

Figures removed due to copyright restrictions.

For calculation of d', consider the % of data of a normal curve

Figures removed due to copyright restrictions.





Response bias c = -0.5[z(H)+z(F)]
C = - 0.5 [z(0.5) + z(0.023)]
C = 1 (bias for saying "no")

Sensitivity d'= z(H)-z(F)
d'= z(0.5) - z(0.023)
d'= 0 – (-2)
d' = 2

In Excel: z(X) is NORMINV(x,0,1) e.g. NORMINV(0.84,0,1)

Conclusion on d’

- **d' is a measure of sensitivity.**
- The larger the d' value, the better your performance.
- A d' value of zero means that you cannot distinguish trials with the target from trials without the target.
- A d' of 4.6 indicates a nearly perfect ability to distinguish between trials that included the target and trials that did not include the target.
- **C is a measure of response bias.**
- A value greater than 0 indicates a conservative bias (a tendency to say `absent' more than `present')
- A value less than 0 indicates a liberal bias.
- Values close to 0 indicate neutral bias.

SDT, Perception and Memory

"Yes-No" paradigms

- A research domain where SDT has been successfully applied is in **the study of memory**.
- Typically in memory experiments, participants are shown a list of words and later asked to make a "yes" or "no" statement as to whether they remember seeing an item before. Alternatively, participants make "old" or "new" responses. The results of the experiment can be portrayed in what is called a **decision matrix**.
- The hit rate is defined as the proportion of "old" responses given for items that are Old and the false alarm rate is the proportion of "old" responses given to items that are New.

	Old	New
Say "Old"	Hit	False Alarm
Say "New"	Miss	Correct Rejection

Costs and Utilities of d’

What are the costs of a false alarm and of a miss for the following:

- A pilot emerges from the fog and estimates whether her position is suitable for landing
- You are screening bags at the airport

Figure removed due to copyright restrictions.

When is a low d’ an advantage ?

Right Face Model for Tiny Displays & Advertising

Figure removed due to copyright restrictions.

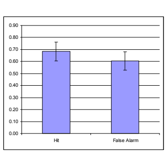
If you can make subtle changes and be sure the person will identify the face as Normal when it has not been manipulated (hit), and when it has been manipulated (false alarm), you have created a condition where the observer cannot see the difference and will have a very low d’.

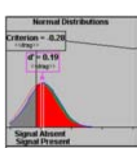
CogLab 1: Signal Detection Results

- d' is a measure of sensitivity.
- A d' value of zero means that ??????
- A d' of 4.6 indicates a nearly perfect ability to distinguish between trials that included the target and trials that did not include the target.
- C is a measure of response bias.
- A value greater than 0 indicates ?????
- a value less than 0 indicates ?????
- Values close to 0 indicate ?????

	Hit	FA	d'	C (bias)
S1	0.33	0.57	-0.60	0.13
S2	0.77	0.37	1.07	-0.19
S3	0.73	0.57	0.46	-0.40
S4	0.83	0.90	-0.31	-1.12
S5	0.70	0.60	0.27	-0.39
S6	0.73	0.63	0.28	-0.48

	Hit	False Alarm	d'	C (bias)
Mean	0.68	0.61	0.19	-0.41





N=6 ☹

Wolfe et al (2005)
Detection of Rare Target


NATURE|Vol 435|26 May 2005

nature

BRIEF COMMUNICATIONS

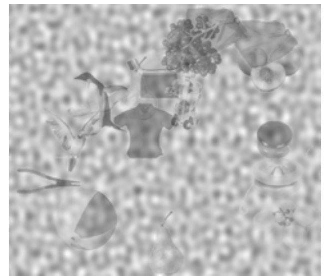
Rare items often missed in visual searches

Errors in spotting key targets soar alarmingly if they appear only infrequently during screening.

“Our society relies on accurate performance in visual screening tasks. These are visual search for rare targets: we show here that target rarity leads to disturbingly Inaccurate performance in target detection”

From Jeremy Wolfe.

What happened?



When tools are present on 50% of trials, observers missed 5-10% of them

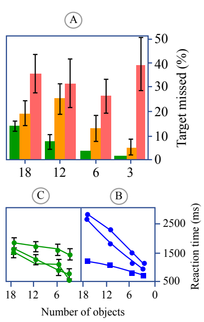
When the same tools are present on just 1% of trials, observers missed 30-40% of them

Find the tool

A problem with <u>performance</u>, not searcher competence.

Slide from Jeremy Wolfe

Here, the important errors are <u>Misses</u>

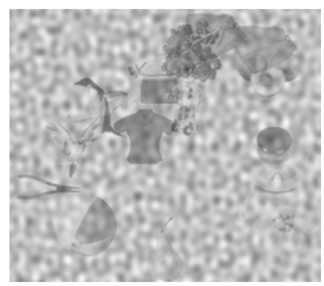


The effects of target prevalence on search performance. A. Error rates for rare targets (red bars, 1% prevalence), less rare targets (orange bars, 10% prevalence) and common targets (green bars, 50% prevalence). Data from 12 observers are averaged. Error bars show ± S.c.m. for those 12 error rates. B. Reaction times for 50% prevalence Typical reaction times are longer when the target is absent (circles) than when targets are present (squares). Diamonds show miss-error reaction times. C, Reaction times for 1% prevalence. At low Prevalence, 'absent' responses are faster than 'present' responses. This leads to increased error rates. Error bars show ± S.e.m.

Figure by MIT OpenCourseWare.

Wolfe, Horowitz, Kenner (2005). Nature

In a nutshell



If you don't find it often,

you often don't find it.

Find the tool

Slide from Jeremy Wolfe

Some implications

- d' measures sensitivity
- Improved sensitivity is only available by providing more information (make the signal easier to detect) or by upgrading the sensory system (make the individual responses more consistent).
- The response criterion can be changed, either deliberately, or by altering perceived costs and utilities (used in social sciences, political sciences)

Some caveats

- Assumptions of equal variance and normal distribution are essential to a simple, one-shot estimation of d'
- *Estimating more points on an ROC curve gives a less biased (assumption-free) estimate*
- Significance tests for differences in d' are available

Map of the course

- I - Signal Detection Theory : Why it is important, interesting and fun?
- II - Some Basic reminders
- III - Signal Detection Theory: Core
- IV – Signal Detection Theory: Advanced

Contingency table

stimulus

response	s0	s1
r0	90	20
r1	10	80
	100	100

Principle of ROC analysis

- ROC = Receiver Operating Characteristic
- or: Relative Operating Characteristic (you choose...)
- Attempts to estimate 2 independent statistics:

- Difference between the means of the two distributions (d')
- The location of the decision criterion (β)

ROC curve

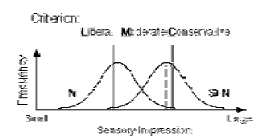
The ROC curve is a graph of HIT rate on the vertical axis and a graph of FA rate on the horizontal axis.

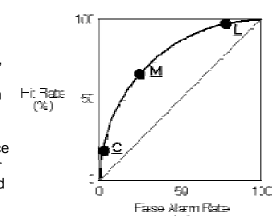
The ROC capture in a single graph the *various criterion*, given a signal and noise distribution.

If the criterion is high (conservative), both the HIT and FA will be high. When the criterion move lower, both HIT and FA decrease.

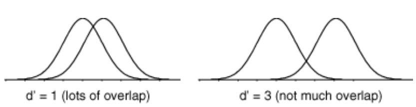
Notice that for any reasonable choice of criterion, HIT rate is always larger than FA, so the ROC curve is bowed upward.

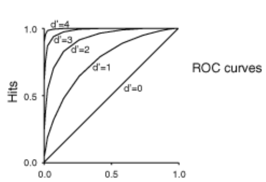
What does the diagonal mean ?





d': how much overlap is there?





Application of ROC curve:

Comparing performances of Machine and Human Vision for a Visual Search task

Figure removed due to copyright restrictions.

Application of ROC curve

Comparing performances of Machine and Human Vision

Figure removed due to copyright restrictions.

9.63 Laboratory in Visual Cognition
Fall 2009