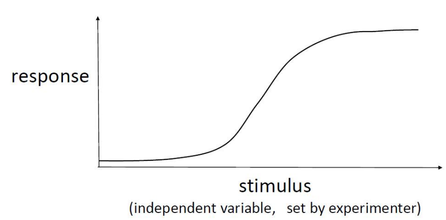
Most of our discussion in this course up to now has been about early vision problems and how the brain solves these problems. We have studied the problems at different levels, from low level neural circuits descriptions up to high level descriptions. In the next few lectures, we will consider how *well* humans solve these problems. These problems wil include basic ones like detecting a change in image intensity or color in a region, or detecting motion a depth increment from disparity, or discriminating the slope of a surface.

The term *psychophysics* refers to experimental methods that measure the mapping from some physical stimulus to a response. A person is shown some images – usually presented on a display screen – and answers questions about the images by pressing on some buttons. We want to characterize how responses depend on the parameters of the images. We are interested in the underlying perceptions rather than the responses themselves. However, we can only find out about the perceptions by asking people to press buttons. (If one is doing psychophysics on monkeys, one can ask them to press buttons and one can also record from cells in their brains. Both kinds of experiments count as psychophysics.)

Psychometric function

A *psychometric function* is a mathematical function from a stimuli level (a parameterized variable) to a response level. The response can be a parameter level that is set by an observer, but for most of our examples it will be a statistic such as percent correct in some task. We will typically consider S-shape (called sigmoid shaped) psychometric functions.



A typical task is "contrast detection". A background patch of intensity I_0 is presented and a central square with an intensity value $I + \Delta I$ is presented, where ΔI is negative or positive. The task could be to judge if there is an increment or decrement. In order to get a psychometric curve that is 'S shaped' and increasing, one could plot the percentage of times that the subject responded that the center was an increment. The response would go from 0 percent (when in fact there is a large decrement) to 100 percent (for a large increment).

In the slides, I discussed a few other ways to set up the problem. One could have a square that is an increment only, and the task would be to say if it is in the left or right half of the display. If the ΔI is very small, then the subject will be at 50 percent correct, and as the ΔI increases, performance will rise from 50 percent to 100 percent.

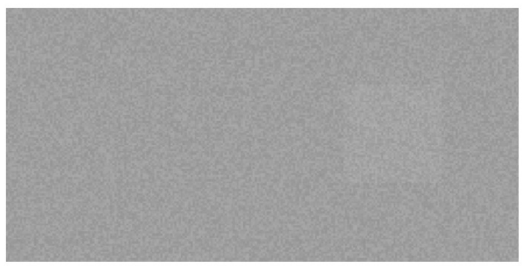
Psychometric curves are typically not step functions. Rather there is a gradual change in performance because of various sources of uncertainty that subjects face when doing the tasks:

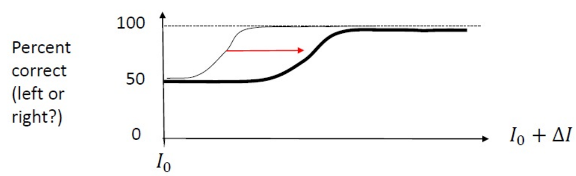
- Noise in the display or stimulus (because it is a physical device)
- Random number generators in the computer program that creates the display image

- Noise in the sensors/brain
- Limited resolution of the display or vision system e.g. finite samples in the photoreceptor grid
- Subjects press the wrong button (stop paying attention)

Different sources of uncertainty play more or less of a role in different experiments.

As the noise or uncertainty increases, it takes a higher stimulus level to reach some performance level like 75 % correct. In the example below, there is a grey square on the right but it is hidden somewhat because of the noise that is added. The more noise that is added, the more difficult it would be to say whether that increment square were on the left or on the right. So, to achieve say 75% correct, one would require a larger ΔI. This would move that 75% value to the first.

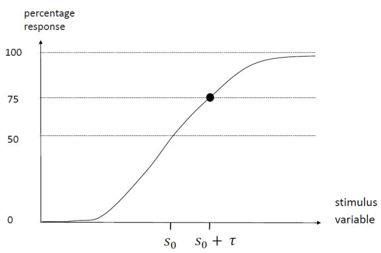


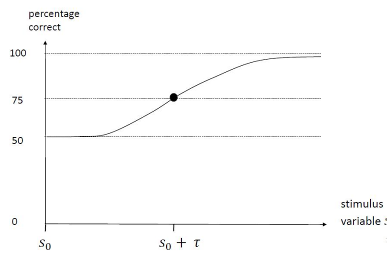


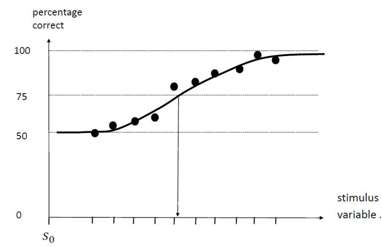
In the figure, I have shown the curve shifting to the right as noise is added. But sometimes the psychometric curve is stretched, or it may be both stretched and shifted. It depends on the task and what is being measured.

Psychophysical thresholds

A psychometric function has a lot of information, and often we just want to summarize it with one number. We arbitrarily take a particular performance level (for example, 75 percent correct) and consider the stimulus level that produces this performance level. This stimulus level is called a *threshold*. Such a threshold can be defined whether the responses go from 0 to 100 percent (left below) as in the case of deciding if an center square has an intensity increment or decrement with respect to the background or from 50 to 100 percent (middle below). The definition depends on the experiment.







In a real experiment, one need to *fit* a psychometric to response data. One then takes the 75 % threshold point *from the fitted curve* rather than from the data. Note that the fit is never perfect (above right), and one has to make assumptions about the family of curves, e.g. `https://en.wikipedia.org/wiki/Logistic_function`. Typically it doesn't matter so much which family of curves one uses, as the exact threshold values are not the main point. Rather, as you will see with some examples, what is more interesting is how the threshold values vary in different versions of the experiment.

Finally, one often thinks of thresholds as values above which a person can do the task and below which the person cannot do the task. (Recall for example Panum's fusion area for binocular stereo vision.) But that is not correct: one's ability to do a task varies *continuously* with the level of the stimulus.

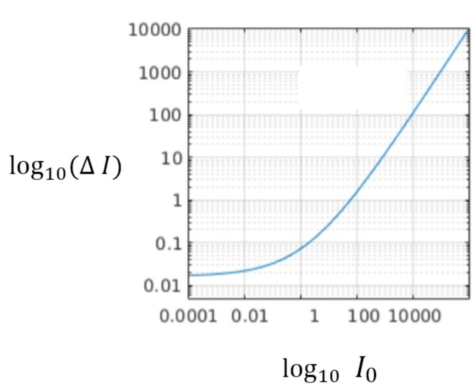
Weber contrast and Michelson contrast

In the examples discussed above, we were considering how well observers can detect a change in intensity ΔI relative to some background intensity I_0. Define the Weber contrast

$$Weber\ contrast \equiv \frac{\Delta I}{I_0}.$$

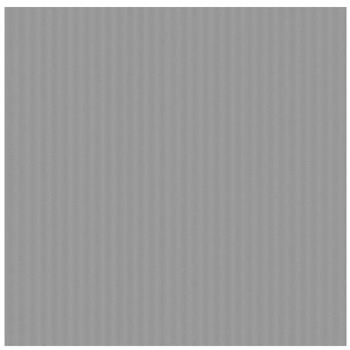
We saw a similar definition of contrast in Assignment 1 where the response of a retinal cell was defined by the difference between the center and a surround, normalized by the mean.

It often happens that the threshold ΔI depends on the I_0. So as I_0 increases, ΔI also increases. One then plots the thresholds ΔI versus the intensity I_0, called a "threshold versus intensity" plot (or TVI). An example is shown below.



The thresholds here are roughly constant when I_0 is low, i.e. detecting the threshold doesn't depend on the background level when the bcakground level is low. But then when the background level is high, the thresholds rise. Here we have a constant rate of increase of the threshold with the background intensity. But note this plot is on a log-log scale, and the slope is about 0.5. Constant slope on a log-log scale gives a power law, where the power is the slope. In this case, we have a square root behavior.

For many experiments in vision, the stimulus is a 2D sinusoid variation. An example is a 2D intensity pattern such as below. The task might be to decide if the 2D sinusoid is vertically or horizontally oriented. In order to make that judgemnt, you need to be able to see the 2D sinusoid! We would like to know how well people can perform this task as a function of the range of intensities in the pattern, and also whether performance depends on the frequency.



When the stimulus for an experiment is a sinusoid, one often defines contrast in a different way than above. Define the Michelson contrast:

$$Michelson\ contrast \equiv \frac{I_{max} - I_{min}}{I_{max} + I_{min}}.$$

Note that this quantity ranges from 0 to 1, where 0 means constant intensity (no contrast) and 1 means maximum contrast. In the example image beow, the Michelson contrast is 0.02.

To understand this definition, write it slightly differently as

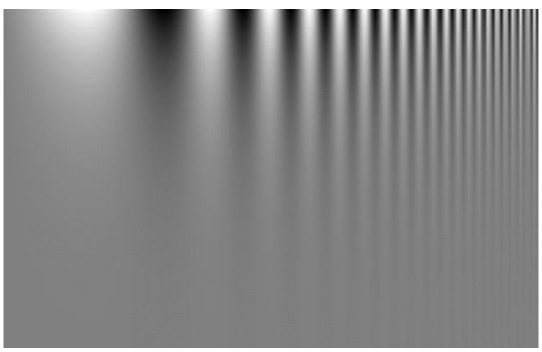
$$\frac{(I_{max} - I_{min})/2}{(I_{max} + I_{min})/2}$$

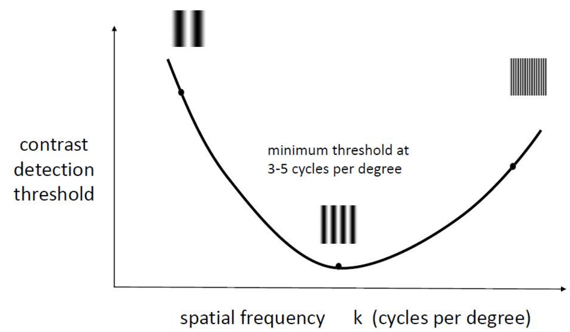
For the case of a 2D sinusoid function, $I(x) = I_0 + a \sin(2\pi k_0 x)$, the numerator is the amplitude a of the sinusoid and the denominator is the mean I_0 of the sinusoid, so the contrast would be a/I_0.

Contrast detection thresholds

2D sinusoidal stimuli are often used in psychophysics to examine sensitivity to oriented structure and structure at different frequencies. Interestingly, contrast thresholds vary significantly with frequency. To illustrate, consider the demo below which shows an image whose spatial frequency varies continuously from left to right and whose contrast increases from bottom to top. Note that the perceived boundary between grey (contrast below threshold) at the bottom and white/black alternation at the top is not a horizontal line, but rather the threshold seems to dip down and up. *The contrast threshold is lowest at the middle frequencies* – i.e. performance is best at the middle frequencies.

A formal experiment to measure contrast detection thresholds at various spatial frequencies would measure thresholds from images that each contain just one spatial frequency. Such a threshold curve is illustrated on the right. Note that the experiment would try to measure a threshold contrast for each frequency k. The curve would then separate the contrasts for which people could get less than 75% correct at the task of judging the orientation (horizontal versus vertical) versus contrasts at which people were more than 75 % correct – namely contrasts below or above the curve respectively.





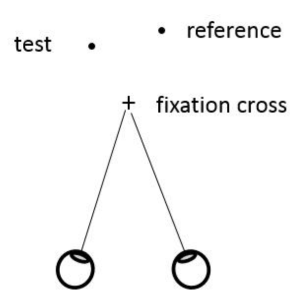
Binocular disparity discrimination

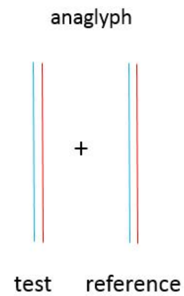
Below I illustrate a common depth discrimination task based on binocular disparity cues. The subject fixates (verges) on a cross, so the cross has disparity 0. A “test” and a “reference” stimulus is also shown, which are vertical lines presented at different depths. In practice these are displayed on a monitor, so that they produce different binocular disparities which give rise to different depth perceptions.

Suppose we hold the disparity of the reference constant, and we vary the disparity of the test, and the task is to say if the test is closer or further than the reference. Responses (“test further”) will go from 0 percent (when test is much closer) to 100 percent (when test is much further). As usual one takes some arbitrary percent correct level (say 75 percent) as the threshold.

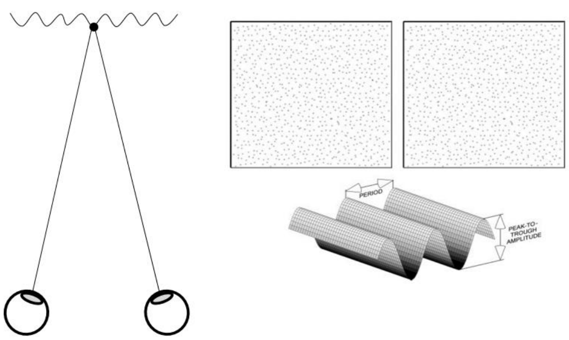
Note that one obtains a different psychometric function for each reference depth, and hence one obtains a different threshold for each reference depth. One can plot the thresholds as a function of reference depth (not shown here or in slides).

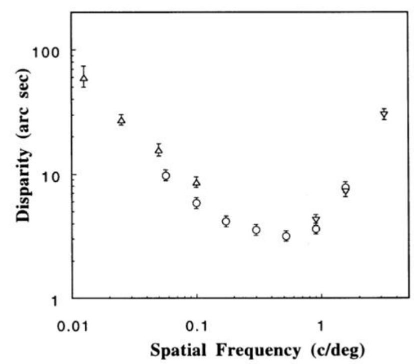
There are various versions of such an experiment. In the slides, I showed a square on background configuration, for a random dot stimuli. The idea is similar. There is reference depth Z_0 which might be the square, and a test depth Z_0 which might be the background (or vice-versa), or this could be expressed as a reference and test disparity.





One can also measure binocular disparity thresholds with 2D sinusoids. One would define a random dot disparity image and the disparity itself would vary as a 2D sinusoid! The figure below which is from a paper by Banks (2004). The data plot on the right shows the threshold amplitude of the disparity disparity of the sinusoid as a function of spatial frequency of the depth variations of teh sinusoid [1]. Note the threshold levels of disparity are remarkably low. A threshold of 5 arc seconds of disparity corresponds to an amazingly small depth amplitude. (See Exercises.)





The lowest thresholds occur at spatial frequencies below 1 cycle per degree of visual angle, which is close to a factor of 10 less than the spatial frequencies where the peak sensitivity for luminance contrast sensitivity occurs (which I mentioned earlier is about 3-5 cycles per degree).

This result is not so surprising, although the reason is rather subtle. The basic idea is that in order for the visual system to be sure that the disparity is varying with some spatial frequency k, it needs to match as many distinct image intensities as possible between the left and the right eye. ("If I only have 10 samples, then I'm not sure, but if I have 10000 samples, then I'm much more sure.) So for a given image pair, the image intensity function needs to vary across space at a much higher rate than the disparity can vary. *What makes this argument especially difficult to understand is to relate random dot images to spatial frequencies. Later in the course, we will see that it is possible to relate the two. But we need to do more work for that.*

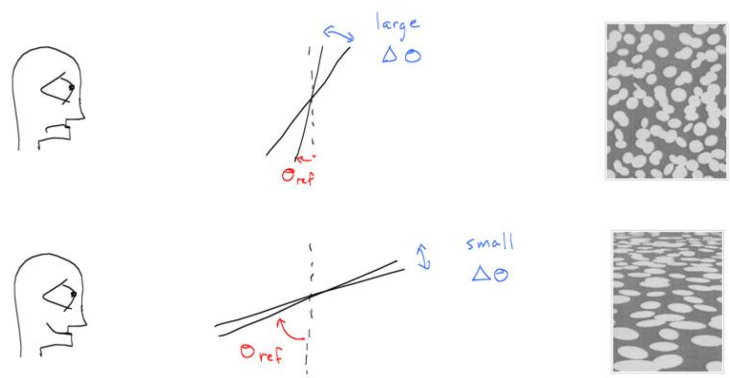
[1] from a paper by Bradshaw and Rogers

Slant from texture

The last example we discuss today is slant from texture. Here the noise is often not pixel noise, but rather it is randomness in the texture pattern itself, namely the shape and size of texture elements. Even if we assume the visual system knows the mathematics of perspective mappings from 3D, it cannot know the size and shape of the texture elements in 3D if these are random and there will be some undercertainty in the surface slant and tilt.

One key finding is that it is inherently more difficult to discriminate the slant of a textured surface when that surface is close to frontoparallel than when it is highly slanted away from the line of sight. Here the task specifically is: given two images (reference and test) of a textured surface, decide which has greater slant. The claim is that thresholds decrease as slant increases.

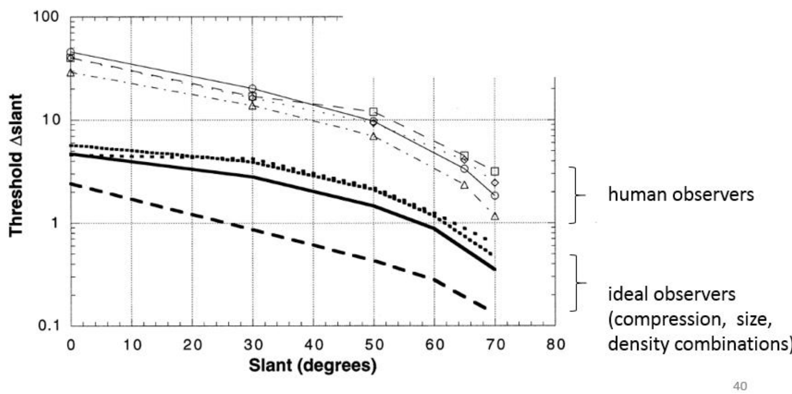
To understand why slant is more difficult to discriminate when a surface is close to frontoparallel, consider the case that the surface is covered with disks (circles). If a disk is slanted slightly away from frontoparallel, this doesn't change the projected shape by much – the aspect ratio (width:height) of the disk in the image is $\cos \sigma$ which remains close to 1 when $\sigma \approx 0$, since the cosine curve is flat at 0. When σ is large, however, the cosine function changes more quickly with σ and so a small change in slant σ leads to a larger change in the aspect ratio and the larger change would allow for greater discrimability of the slant of the disk.



The experiments illustrated in the figure above don't use disks. Instead they use ellipses which are dropped down on the plane in random positions and orientations. These ellipses have a random distribution of sizes and aspect ratios. (A different example texture was shown in the lecture slides.) But the same idea holds, namely that there is relatively less information about the foreshortening of the ellipses when the surface is frontoparallel than when it is slanted. To estimate slant, the visual system needs to use probabilities of various ellipses.

Consider a computational model in which the observer *knows* the probability distribution of the texture elements (ellipsoids) and it knows about the geometry of perspective. Suppose this model observer were given an image of a surface that is slanted in depth, and tries to estimate the slant. Because there is randomness in the texture element distribution, its estimate of the slant would not be exact. The question is, *would such a model observer also have different thresholds for different reference slants* ? The answer is, yes! Indeed it was shown about 20 years ago in a very nice set of experiments[2] that these model observers do better than humans (which is not surprising, since the model observers know every there is to know about the situation) but that their performance varied with refererence slant in a similar pattern to human performance.

There are a lot of details to be specified here, and I am omitting most of them. Different model observers were tested using combinations of the texture cues to slant, namely foreshortening (also called "compression"), size, and density. Never mind the details for now. The main point, which you can see in the figure, is that the threshold on slant decreases as the slant increases: we are better at discriminating the slant (angle) of highly slanted surfaces than frontoparallel surfaces. This is true both for ideal observers and for human observers. So, humans are just using the information that is there.



To summarize some of the main ideas of today: plots of thresholds as a function of scene parameters can reveal two different aspects of how the visual system is solving a problem. First, the plots can reveal underlying mechanisms. e.g. There may be cells that encode *limited* ranges of size, orientation, disparity, etc. The contrast threshold plots from today are a good example. Note that for these types of examples, the performance of human observers might exhibit quite different patterns than the performance of model observers; people might not be using information that *is* available for whatever reason. Second, the plots can reveal how the inherent difficulty of the computational problem varies over different ranges of parameters. Slant from texture is a good example of this. For such examples, human and ideal observer performance tends to exhibit similar patterns, with the humans performing consistently worse than the model (since we humans have internal "noise").

[2] by David Knill and others