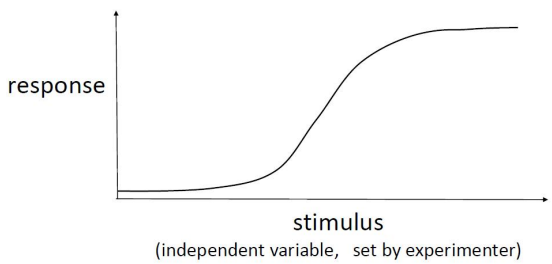
Most of our discussion in this course up to now has been about early vision problems and the early processing in the brain to solve these problems. In the next few lectures, we will turn to how well humans solve these computational problems. These problems may include detecting a change in image intensity 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. Psychophysics is the field of science that characterizes how responses depend on the parameters of the images. One is more interested in the underlying perceptions, and less interested in the responses themselves. But typically 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, or it can be a statistic such as percent correct in some task. Most of our examples will consider the latter. We will typically consider S-shape (called sigmoid shaped) psychometric functions.



An example task a background patch of intensity I_0 and a central square with a different value of intensity $I + \Delta I$, 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 would plot the percentage of times that the subject responded that the center was an increment. The response would go from 0 percent (for large decrements) to 100 percent (for large increments).

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. But as the ΔI increases, performance will rise from 50 percent to 100 percent.

Psychometric curves are typically not step functions. The reason that there is a gradual change in performance is that there are 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

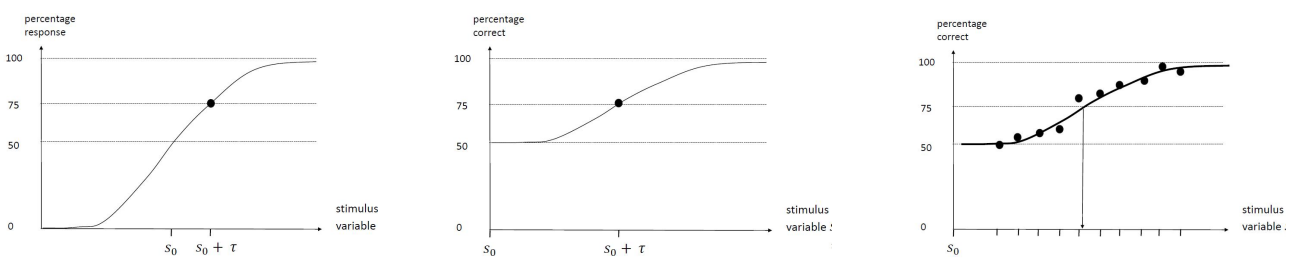
- Noise in the sensors/brain
- Limited resolution of the display or vision system e.g. finite samples in the photoreceptor grid
- Subjects press the wrong button (stop paying attention)

Different sources of uncertainty play more or less of a role in different experiments. In general, as the noise or uncertainty increases, it takes more stimulus to reach high performance. Sometimes the psychometric curves stretches out as uncertainty (noise) increases and sometimes it shifts the right, and sometimes it both stretches and shifts.

It is important to note that some of the factors that limit performance are within the observer (noise in the brain, failing to pay attention) but that some factors are inherent in the stimulus. Even a vision system that had no 'brain noise' and always paid 100 percent attention would still make mistakes since the stimulus itself could have randomness – so even an *ideal observer* would need to guess sometimes.

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 in the case of a psychometric function going from 50 to 100 percent (middle below) as in the case of detecting if an increment is present (e.g. s_0 is the background intensity of square).

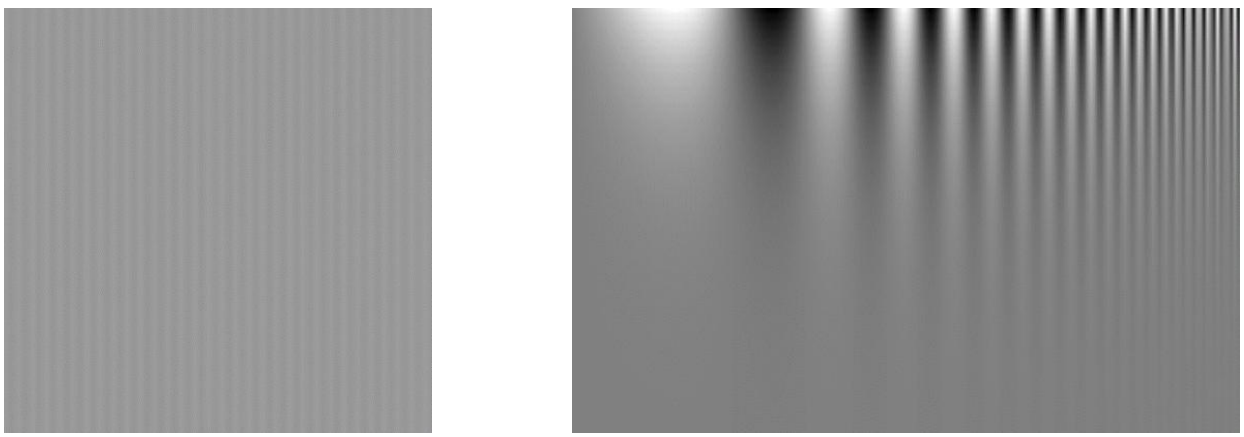


In a real experiment, one fits the parameters of some model curve to some noisy data. One then takes the 75 % threshold point *from the fitted curve* rather than from the data. Note that the fit is never perfect (above right), and often one makes strong assumptions about the shape of the curve. This is ok, 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 values vary in different compare threshold values for different versions of the experiment – that is, across different psychometric curves. This will be more clear once you see a few examples.

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 of course that's oversimplified, since one's ability to do a task varies continuously with the amount of stimulus relative to the noise.

Michelson contrast

For several of the examples that we discuss, the stimulus is a 2D sinusoid variation. An example is a 2D intensity pattern such as below on the left. The task might be to decide if the 2D sinusoid is vertically or horizontally oriented. 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.



Define the Michelson contrast:

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

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, $I(x) = I_0 + a\sin(2\pi k_x x)$, 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.

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 above on the left, the Michelson contrast is 0.02.

Contrast detection thresholds

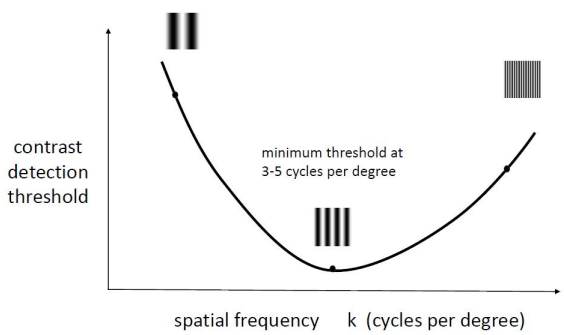
2D sinusoidal stimuli are often used in psychophysics to examine sensitivity to oriented structure and structure at different scales. Consider the example above right, which shows an image whose spatial frequency varies continuously from left to right and whose contrast increases from bottom to top. 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.*

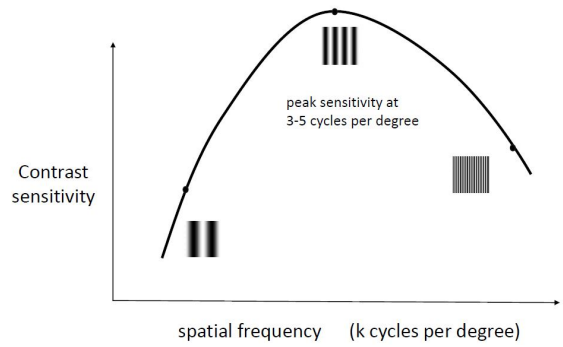
The figure just mentioned is a demo, not an experiment. A formal experiment to measure contrast detection thresholds at various spatial frequencies would measure thresholds from images

that each contain just one spatial frequency, such as above on the left. See the Exercises for some examples.

One often plots *contrast sensitivity* which is defined as the inverse of contrast. For example, a contrast detection threshold of 0.02 is equivalent to a contrast sensitivity of 50. Just as contrast threshold curves typically have a minimum at some middle frequency, contrast sensitivity curves have a maximum at that same middle frequency. See exercises for some examples.

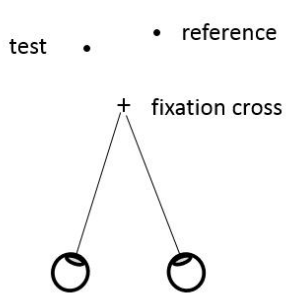
The shape of the contrast sensitivity curves is presumably due to constraints on how many cells in the visual system have their peak sensitivity to different spatial frequencies. We are more sensitive to those spatial frequencies to which more of our cells are tuned. (Recall that DOG cells in the retina and LGN each have a particular range of sizes and this range varies from the fovea to the periphery. As you will see in one of the exercises, contrast sensitivity varies with eccentricity as well.)

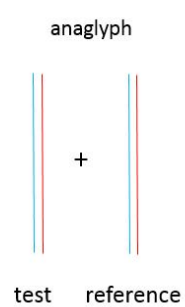




Binocular disparity discrimination

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.



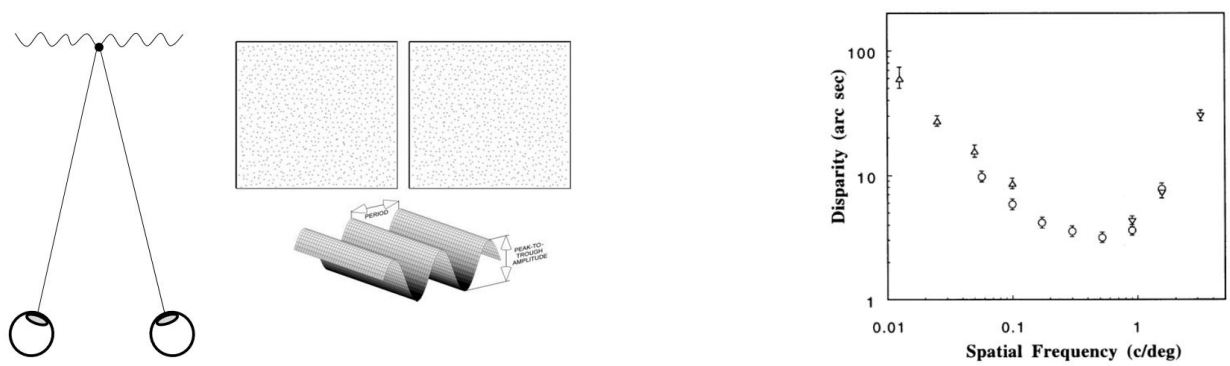


Suppose we hold the disparity of the reference constant, and we vary the disparity of the test, and the task is to say if the test is closer or further than the reference. Responses ("test further") will

go from 0 percent (when test is much closer) to 100 percent (when test is much further). As usual one takes some arbitrary 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).

There are various versions of such an experiment. In the slides, I showed a square on background configuration. The idea is similar. There is reference depth which might be the square, and a test depth which might be the background (or vice-versa). 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! Since figure below which is from a paper by Banks (2004). The data plot on the right shows the threshold amplitude of the disparity sinusoid as a function of spatial frequency[1]. Note the threshold levels of disparity are remarkably low. A threshold of 5 arc seconds of disparity corresponds to quite a small depth amplitude. (See Exercises.)



Also note that 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).

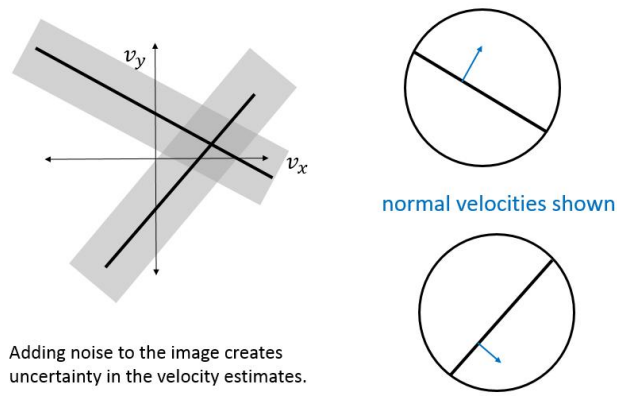
Is this surprising? Not really. For your visual system to measure small spatial *variations* or gradients in disparity from random dot stereograms, it needs to precisely represent the disparities in those regions (since otherwise, how would it know that the disparity is changing?) But to precisely estimate the disparity of a small local region – i.e. to match small local regions of the left and right image – the visual system requires that the image has intensity gradients that are high e.g. sharp edges or lines *and* that the visual system needs to be sensitive to these sharp image structures. As we will see a few lectures from now, sharp edges and lines in the intensity mean that there are high spatial frequencies in intensity present. (Wait for after study break for this.)

Motion Discrimination

Now that we know what noise is for, let's consider how we can add noise to other cues. For motion, if we add pixel noise to a video $I(x, y, t)$, then the partial derivatives of I with respect to x, y, t

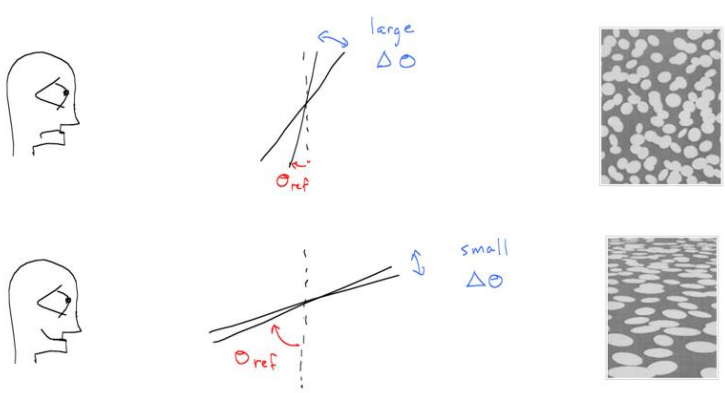
[1] from a paper by Bradshaw and Rogers

essentially become noisy.[2] The result is that the motion constraint line would become uncertain. There would be a *distribution* of motion constraint lines. The intersection of constraints would then not give a unique solution but rather would give a *region* of uncertainty for the solutions. Experiments have shown that this is indeed what happens perceptually. Intuitively, it is not surprising. Adding pixel noise makes it difficult to judge exactly what the image velocity is.



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. We will discuss this more next lecture, but let me just mention one important idea here, illustrated below.



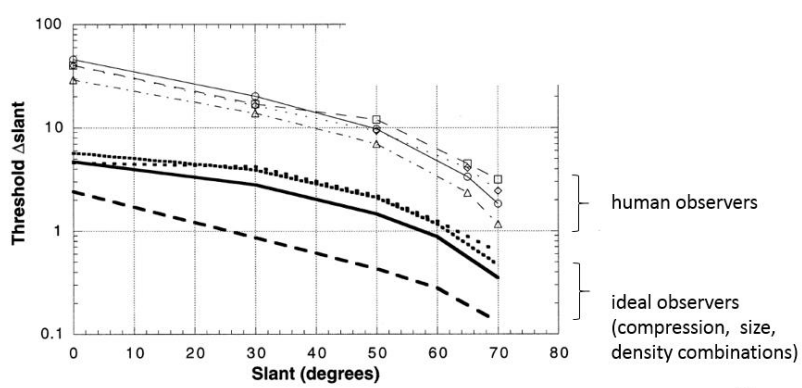
The claim is thatit is inherently more difficult to discriminate the slant of a textured surface when that surface is close to frontoparallel than when it is highly slanted away from the line of

[2] We can't take derivatives of a discrete image, but we can take local differences at neighbhoring pixels, and they would be noisy.

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, notice that slanting a (circular) disk slightly away from frontoparallel 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. 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. The calculation is non-trivial for an ideal observer, and not surprisingly, the visual system does not perform as well as various ideal observers. (The figure below is from Knill 1998 and shows a few different ideal observers that were used to estimate the slant. These different ideal observers used 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, 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. There may be cells that encode *limited* ranges of size, orientation, disparity, motion, etc. The contrast sensitivity 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 ideal 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 human typical performing consistently worse than the ideal (since humans are not ideal).