23.2: How Many bits/min² Are Needed for the Perfect Display?

Stanley A. Klein, Thom Carney
University of California at Berkeley, Berkeley, CA

How many bits/min² are needed for the perfect display?

Stanley A. Klein and Thom Carney
School of Optometry, U.C. Berkeley
Berkeley, CA 94720

How many bits per square minute should a monitor be capable of displaying? The answer, of course, is ambiguous in that it depends on the task. One answer may be appropriate for displaying x-rays needed in a medical diagnosis and another appropriate for reading text on a word processor. We can make the answer unambiguous and unique by asking for a “perfect” monochromatic display. By perfect we mean a display that is able to present an arbitrary image such that an observer can not distinguish it from the original. This criterion of being perceptually lossless is much more demanding than is generally realized. The computer monitor being used to write this article is displaying about .2 bits/min². We will show that the human visual system requires the monitor to be capable of about 160 bits/min² for a perceptually lossless display.

We choose to quantify the information capability of a display in units of bits/min² because that is the quantity relevant to the observer. Bits/min² is the product of bits/pixel times pixels/min². The Macintosh SE display I am viewing right now has a pixel size of about .3 mm. At my 24 cm viewing distance each pixel subtends about 2.5 min so each pixel has an area of 6.25 min². This display only has one bit/pixel so N, the total number of bits/min², is given by: N = 46.25 x .16 bits/min². For image processing, on the other hand, a common display is 256 x 256 pixels viewed from 8 times the screen size (Grod, 1988; Watson, 1987) which gives about 1.68 min²/px (2.82 min²/px). If each pixel could display 8 bits of luminance (256 grey levels) then the display would be capable of N = 46.25 x 8.16 bits/min².

The last example is standard in the image processing community. Such a display, however, is not capable of showing the full range of images that can be discriminated by the visual system. For example, a 50 c/deg square wave grating could not be displayed because of the coarse pixels, nor could a near-threshold 4 c/deg grating be shown because of the coarse grey levels. The usual calculation of the human visual system capability goes as follows. Since the highest spatial frequency seen by the visual system is about 60 c/deg, the finest sampling that is needed is .5 min. So the smallest pixel size is taken to be .5 min or .25 min²/pixel. The number of luminance levels is a bit harder to crudely estimate since it depends on several factors. A naive guess might be that about 10 bits should be sufficient to display all discriminable grey levels. If so then a total of N = 10/.25 = .40 bits/min² would be estimated as the needed capability of a display. It is the purpose of this paper to show that this calculation is wrong and that rather than 4 pixels/min², 9 pixels/min² are needed, and instead of 10 bits of luminance, about 18 bits are needed, leading to a total of about 160 bits/min² needed for the perfect monochromatic display.

Blur discrimination and the maximum pixel size

The visual system’s resolution ability places an upper limit on the separation between pixels. A resolution task involves the discrimination of two lines from a single line. It might be thought that a high acuity task places an even stronger limit on the pixel size. Consider the problem of displaying a very thin line that is slightly tilted away from vertical. If the line is always one pixel wide then the line tilt would be accomplished by a succession of abrupt one pixel jumps. Since the vernier acuity threshold is about 5 sec of arc, a pixel size of 3 sec would be needed to eliminate the visibility of the jumps. A larger pixel size would suffice if a “distorting” technique were used to smooth out the jumps. The tilted line would then be two pixels wide whenever the center of the line did not coincide with the center of a pixel. By this device the line’s centroid could be shifted by adjusting the contrast of the two constituent lines. For the tilted thin line, the limitation on pixel size is not the vernier threshold, but rather a resolution threshold since the tilted thin line alternates between being one pixel thick and two pixels thick.

Two methods can be used to determine the pixel size based on resolution data: one theoretical and one experimental. The theoretical argument, based on Nyquist sampling, leads to a .5 min pixel. The argument is simply that since the cutoff spatial frequency of the visual system is about 60 c/deg, the Nyquist theorem seems to imply that with a .5 min sampling any general image should be reproducible. It thus came as a surprise to us that our experiments on the resolution of thin lines (Levi & Klein 1990; Carney, Klein & Levi 1988) implied that a .33 min pixel was required.

The finding that two lines can be discriminated from a single line when they are separated by .33 min is based on our recent experiments (Carney et al., 1988) performed using a VENUS stimulus generator and a Tektronix 605X monitor. The mean luminance was 100 cd/m² and the pixel size was .31 min. Resolution (blur detection) was measured using the method of constant stimuli in which on different trials single lines were intermixed with line pairs of different separations (our method of producing the different stimuli will be discussed next). The resolution threshold was taken at d' = 1. Under these conditions two lines are discriminable from one line based on a blur (width) cue well before they are seen as separate. Normally, resolution involves a task in which the separation between two lines is increased until the pair appears doubled. We adopt a looser definition for the word resolution and allow the observer to use any cue (such as blur or line width) that is available.

The stimulus in our experiments was a positive contrast line four pixels wide. The intensity profile of the four pixels was symmetric and can be written as a(1/2-a) L2-a where each line intensity is specified in units of % min. This is the unit used since line strength depends on the product of line contrast (%) times line width (min). Two moments of this intensity profile are of interest: the zeroth moment (called the line moment) in units of % min:

\[ I = \int P(x) \, dx \]

\[ = a + (L2 - a) + (L2 - a) + a \]
In both panels, the abscissa is the line moment. In the upper panel, the ordinate is the threshold quadrupole moment. The two arrows at the ordinate indicate the quadrupole detection threshold for the two observers. It is seen that the line pedestal facilitates the quadrupole detection.

The lower panel converts the abscissa (L) and ordinate (Q) to the equivalent separation, d, using Eq. 3. The separation, d, is the resolution threshold. It is seen that the resolution threshold decreases from about 0.1 min at low pedestal strengths to about 0.33 min and 0.4 min at higher pedestals for the two observers.

Similar results were found by Levi and Klein (1990) who measured 2-line resolution under similar conditions. One difference in methodology was that Levi and Klein jittered the contrast of the lines by up to 30% from trial to trial, to remove any possible contrast cue that could be used to judge separation. Even with the contrast cue removed, Levi and Klein found resolution thresholds of 0.33 and 0.4 min for 2 observers, when the line pair was about 10 times detection threshold. At larger stimulus strengths the resolution threshold would be further reduced. For our present purpose it isn't necessary to remove contrast cues since we would like our hypothetical tilted straight line to appear to be of uniform contrast.

This finding, that an observer can discriminate a single line from a line pair when the separation is as small as 0.33 min, is at first surprising since it implies that in order to display a thin line a pixel size of 0.33 min must be used. This is smaller than the 0.5 min that is implied by the Nyquist limit based on the visual system's not passing any spatial frequencies above 60 c/deg. The resolution of this paradox is that the line spread function blurs the 0.33 min line so that the retinal sampling is just fine. Thus 0.5 min cone spacing is just fine. The Nyquist limit applies to the band-limited retinal image not the original image on the display. This point can be seen in a simpler way by considering a square wave 60 c/deg grating. This grating can be produced on a display with 0.5 min samples (assuming throughout that the pixel luminance profile is a nice rectangular shape 1 pixel wide). The contrast of the fundamental is 127% (4/0.5). Suppose the optical transfer function at 60 c/deg is 0.02. The retinal contrast would then be 2.5%. This retinal contrast might be just at threshold. Suppose now we used a display in which the pixel size was 0.25 min. Then a rectangular 60 c/deg grating alternating between one white and 3 black pixels would give a display contrast of 8/(3*2) = 190%. This would produce a retinal contrast of 3.6%, which would be much more visible to the observer. This argument shows that it is necessary for the display to be sampled above the retinal Nyquist frequency for a lossless display.

The above calculation of the maximum pixel size points out a recurring theme of our approach. The stimuli are selected to isolate the cue of interest. If the resolution of dots rather than lines had been used to determine the pixel size then a larger pixel would have been judged to be sufficient. With the larger pixel size, however, a tilted thin line would appear blurred at certain points and our goal of having high fidelity for all images would be violated.
Contrast discrimination and the number of grey levels

It is easy to underestimate the number of grey levels needed by a display. Blackwell (1946), for example, found that a 2% luminance change was needed for detection. That value, however, underestimates the capabilities of the human visual system by an order of magnitude. Rather than using disks, one should use a grating between 2 and 6 c/deg, near the peak of the CSF. Campbell and Robson showed that the visual system's sensitivity to square wave gratings was slightly greater than 500, corresponding to a Michelson contrast of 2% or a luminance change of 4%. One can do even better. Nachmias and Sansbury (1974), and Stromeyer and Klein (1974) showed that contrast discrimination can be two to three times better than contrast detection. That is, although the Michelson detection contrast is 2%, the contrast discrimination threshold is lower. A 1% contrast change corresponding to a 2% luminance change (peak to trough) is detectable when a pedestal is present. In order to maintain 2% luminance steps over a 100 fold luminance range (say 2-200 cd/m2), the number of steps, _N_, is given by:

\[ 1.002^N = 100 \]

or \[ N = \ln(100)/\ln(1.002) = 2,302. \]

A logarithmic attenuator with slightly more than 11 bits (11.2) would be needed to produce the 2,302 levels.

Eleven bits of grey levels sounds like a lot but it is an underestimate. The stimuli we have considered so far are static. In the presence of slow image motion, the threshold can be reduced by at least another factor of two. Thus the contrast of a squarewave drifting grating should be discriminable at a level of about .1% in the facilitation regime. Since a contrast discrimination of .1% is half the value expected for static images the number of required levels is expected to be

\[ N = \ln(100)/\ln(1.00) = 4,607 \]

corresponding to \( N = \ln(4.607)/\ln(2) = 12.2 \) bits for a logarithmic digital to analog converter (DAC) to cover the luminance range that extends from 2 to 200 cd/m2.

In the above computations, we have assumed that video DACs are logarithmic, meaning that the input is the logarithm of the intended output. This unfortunately is not the case. Video DACs are linear. The most difficult image for a linear DAC to display is a moving square wave at the peak of the CSF, at the lowest luminance (2 cd/m2 in our example). The visual sensitivity at 2 cd/m2 is not much lower than the sensitivity at higher luminances. This is because at the relatively low spatial frequency of the CSF peak, Weber's law is expected to hold rather than the DeVries-Rose square root law that obtains at higher spatial frequencies. To be conservative, we assume that a moving square wave grating with a .5% luminance change can be detected at 2 cd/m2. In order to span the 100-fold luminance range from 2 to 200 cd/m2 the number of luminance steps is \( N = 100/0.005 = 20,000 \). The number of bits is \( \log(20,000)/\log(2) = 14.3 \) bits.

One might worry that present video boards are unable to achieve 14.3 bits of luminance levels. Not so. The VENUS display that we use for most of our psychophysics experiments puts out low quality bits on each of three colors. Pelli and Zhang (1990) showed how to combine the three outputs using a passive resistor network with all input and output impedances matched. The single luminance output will have about 6 extra bits (the number of extra bits is about half the number of bits of each DAC). This scheme brings us up to 18 bits per pixel. If color is desired then one must use three video boards to get three colors with 18 bits per color.

Total bits/min²

Resolution data implies that the pixel size should not be greater than .33 min. Thus there must be at least nine pixels per min². Contrast discrimination data implies that 14.3 bits per pixel are needed to represent luminance for arithmetic DACs. Thus for a perceptually lossless monochromatic display, about 130 bits per min² are needed. One's first reaction is that this is an absurdly high number. However, the VENUS display that is used in most of our experiments has a .31 min pixel size and 12 bits of output corresponding to 125 bits/min². We are not using the passive resistance network (Pelli and Zhang, 1990) described above, so we must avoid low contrast square waves at the lowest luminances. The limitation of our display is that since it only has 256 x 256 pixels the display subtends only 1.3 x 1.3 deg of visual angle. When we need a larger field we move closer than the 4 m standard viewing distance, but then we are unable to do our resolution experiments.

Effect of compression: "dithering"

By introducing "intelligence" to the display the number of bits/pixel can be reduced. This takes us to the topic of image compression. It is useful to distinguish between three types of compression: "dithering," "visual or CSF models," and "redundancy models." Dithering or "half-toning" is a scheme in which the optical blur of the human eye blends adjacent pixels to obtain intermediate luminance levels. Dithering is especially relevant to the present paper since it doesn't require any intelligence on the part of the monitor. The sender encodes the image with fewer than the original number of luminance levels. The dithering is not detected because the human visual system is insensitive to small luminance changes at high spatial frequencies. If the dithering scheme involves 9 pixels (a 1 min square block), then a nine-fold increase in the number of intermediate luminance steps would be possible. This procedure would lessen the number of grey levels needed by a factor of 9, reducing the number of bits by \( \log(9)/\log(2) = 3.17 \) bits/pixel. So, instead of 14.3 bits/pixel, the 9-pixel dithering reduces the need to 11.1 bits/pixel or 100 bits/min². Although 100 is a sizeable reduction from 130 bits/min², it is still a high number. In any case, one can not count upon the received image to be encoded in a dithered form. The dithering process takes time, and for this reason it is not commonly done. If it is desired to have a display that subtends about 15 deg (a viewing distance about 4 times the display width) then there would have to be 2700 pixels across the display (assuming .3 min pixels). The full display would have more than 5 million pixels, and thus dithering would be quite time consuming. It is in the spirit of the present paper to not
make assumptions about the image, but rather to assume an arbitrary image, in which case we are back to 120 bits/min².

**Visual or "CSF" compression**

A discussion of how many bits/min² are needed to display an image should at least mention the possibility of true image compression taking into account the large savings in bits/min² due to the limited visual processing of the human observer. There are many image encoding schemes that are based on the fact that at spatial frequencies above 40 cpd, the human visual system can discriminate only a very limited number of contrasts. In a recent paper Klein (1990) examines several variants of the Discrete Cosine Transform (DCT) in great detail. When the DCT compression scheme makes appropriate use of the characteristics of the human visual system, it is found that about 20 bits/min² are needed when a logarithmic (Weber's law) quantization is used and about 31 bits/min² are needed for uniform quantization. Although these numbers are drastically smaller than the 120 bits/min² needed for the perfect display, they are, nevertheless, much larger than the 1 or so bits/min² typically found by other image compressors who start with degraded images.

**Redundancy compression**

The big savings in image compression can be found in making use of the great redundancy in many images. A typical image has many large regions of totally uniformity. These regions can be detected and encoded with few bits. The point we would like to make on this topic is to suggest that all studies of image compression should report two numbers: visual system compression and redundancy compression. The advantage in having the first number is that it will help clarify what assumptions about the human observer are being made. The advantage of having both numbers is that one can then see the gain made by the redundancy in the image. The reduction in bits/min² due to redundancy is very image dependent, so it can not be relied on for all images.

**Conclusions**

A monitor must be capable of generating 120 bits/min² in order to guarantee that an arbitrary image can not be discriminated from the original. This result is larger than might have been expected and suggests the visual system can be very demanding. The mismatch between the performance of the visual system under optimal conditions and under the conditions normally used in judging images may account for the difficulty that researchers have had in finding a satisfactory method for judging display performance; image quality and for evaluating compression schemes.

**References**


**Note:** The manuscript by Klein and Carney, "How many bits/min² are needed for perfect image quality?" recently submitted to JOSA has an improved discussion of the issues covered in the present paper.