Properties of the Human Visual System Relevant to Video Coding

Dr. Noel Murphy

September 22, 2004

1 Introduction

Since the first movies, and particularly during the implementation of the first television systems, the psychophysical properties of the human visual system have been used to determine the requirements for comfortably presenting visual material to viewers. The need to use lossy coding, but to minimize the impact of this on the quality of material was central to the video coding standardisation efforts in the last two decades. Here we introduce some anatomical, physiological and psychophysical properties of the human visual system that are relevant to video coding.

2 The Eye

2.1 Overview

The retina of the human eye contains two main classes of photoreceptors, rods and cones. The approximately 100 million cones are responsible for night vision and are essentially monochromatic. The approximately 6.5 million cones are responsible for daylight vision and come in three varieties which are sensitive to different wavelengths, giving us colour vision. The receptors are not uniformly distributed on the retina: the cones in particular are very densely packed approximately in the centre of the retina in a 1°of arc wide region called the *fovea.* The packing in the fovea is a fairly regular hexagonal (triangular) array with about 120 cones per degree of arc. Outside the fovea, the cone spacing falls off rapidly and has a fairly random spacing lattice.

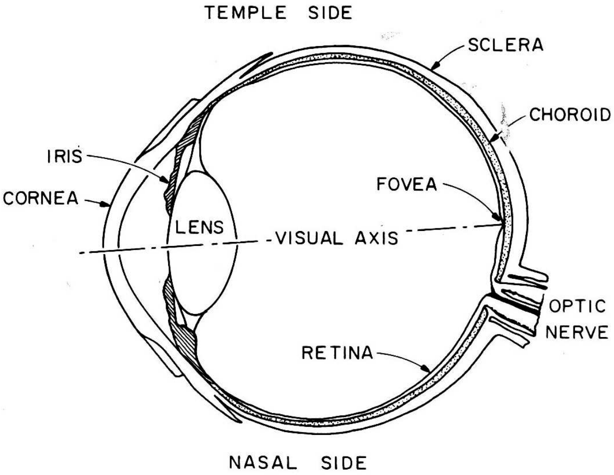


Figure 1: Schematic vertical cross-section of the human eye.

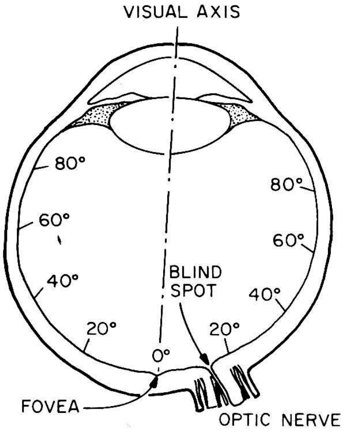


Figure 2: Schematic horizontal cross-section of the human eye.

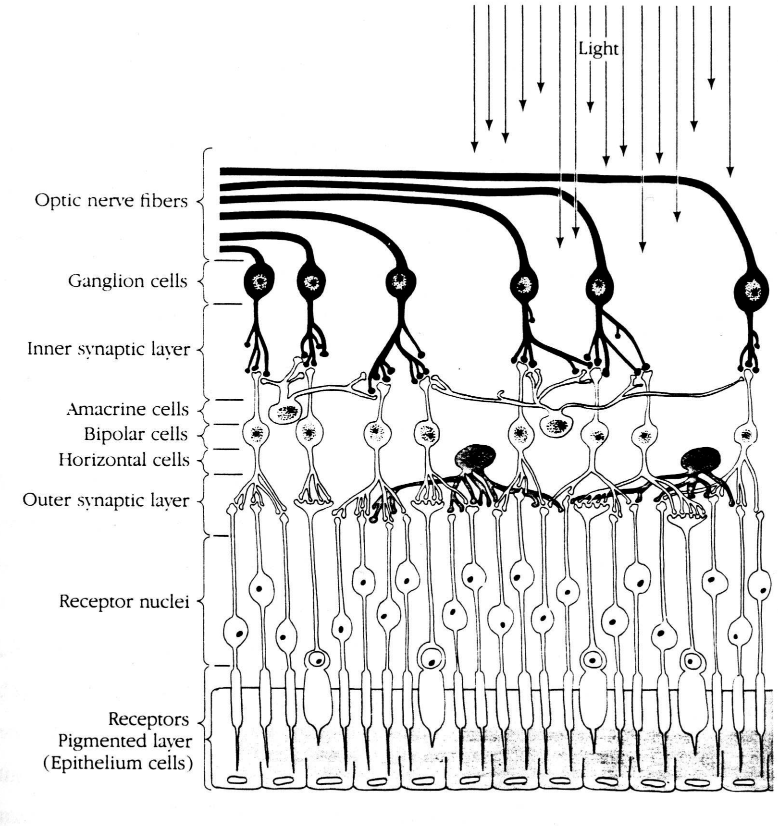


Figure 3: Schematic cross-section through the human retina. Note that the light passes through the semi-transparent layers of neurons before being detected by the photoreceptors.

2.2 The Fovea and Foveal Acuity

The *pupil* is the circular aperture at the front of the eye. Because of diffraction at the pupil and aberrations in the optics of the eye, the eye acts as a low-pass filter which for green light has an upper cutoff of approximately 60 cycles per degree. This means that higher spatial frequencies above the Nyquist limit which would cause aliasing or Moiré effects on the regular foveal array are eliminated.

Under optimal conditions the point spread function has a diameter of about 1 min. of arc. In terms of spatial frequency, the contrast of a retinal image falls by an order of magnitude between 0 and 35 cycles per degree and is negligible above 60 cycles per degree. In the fovea, the blurred retinal image is sampled by an array of cones which form an accurate triangular lattice. The cone inner segments (the active sensing element) fill most of the area between them. This normally has the dual effect of maximizing the quantum catch, (thereby minimizing the photon noise), but also of reducing the spatial frequency response because of integration of light across the cone aperture. However, with a point spread function of the eye of about $5\mu m$, the $2.3\mu m$ diameter of the cone aperture means that the blurring due to integration across the aperture is small. (Loss of contrast is less than 25% at 60 cycles per degree). The minimum centre-to-centre spacing between human foveal cones is about $2.8-3.0\mu m$. With a row spacing in the triangular lattice of $2.4 - 2.6\mu m$ (0.5-0.54 min. of arc) the spatial sampling frequency yields a Nyquist limit of 56-60c ycles per degree. These measurements have been confirmed by the generation of fine laser interference fringes directly on the retina of human observers who report the Moiré patterns formed by the cone mosaic. The Moiré patterns are coarsest at a frequency of 110-120cycles per degree and can be seen up to frequencies of 150-160 cycles per degree. They also show a 60° rotational periodicity and some distortion consistent with a slightly irregular triangular lattice. The fact the Moiré fringes cannot be seen in the fovea under normal viewing condition indicates that the eye's optics remove the higher spatial frequencies from the retinal images which would cause aliasing.

This would suggest that diffraction and aberration are the limiting factors on spatial acuity rather than receptor density. The particular form of the eye represents a choice between a larger pupil diameter, where diffraction would be reduced at the expense of aberrations, or a smaller diameter which would reduce aberrations but cause diffraction to limit acuity. Interestingly, it may actually be the cones' size and spacing which are the ultimate limit to spatial acuity, with the optics evolving to a stage where they do not interfere with the acuity achievable with the minimum cone spacing. Cones

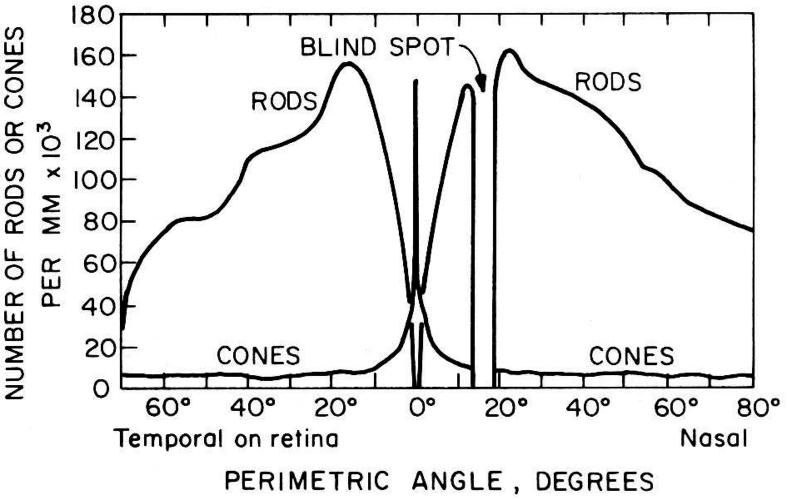


Figure 4: Rod and cone distribution as a function of angular position on the retina.

smaller that $2\mu m$ have never been found, even in very small eyes, and there is remarkable constancy in cone diameter over a large range of eye sizes. The apparent minimum "allowed" spacing of cones, may be required to prevent optical "cross-talk" between the outer segments of receptors. These segments are essentially short optical waveguides, with limited ability to retain incident light due to the limited refractive index differences in the biochemical materials available. Photons captured by one cone could easily cause a photopolymerization (the first step in the detection of a photon) in a neighbouring cone, if the spacing were sufficiently small. Many structural details of the eye, including the pupil diameter and optical quality, may be a consequence of the need to optically isolate cone outer segments, rather than the cone spacing being a consequence of the lack of further evolutionary pressure from the poor optical quality. The match between the highest frequency passed by the eye's optics and the resolution limit set by the spacing between cones at the foveal centre has long been known. This recent explanation of which of the two factors, optics or cone separation, is responsible for the actual acuity value puts the meaning of acuity in new light.

2.3 Colour Transduction

Also interesting is the way the retina has adapted to accommodate colour vision. The retina contains three subpopulations of cones, each sensitive to incident light with frequencies in one of three particular bands in the

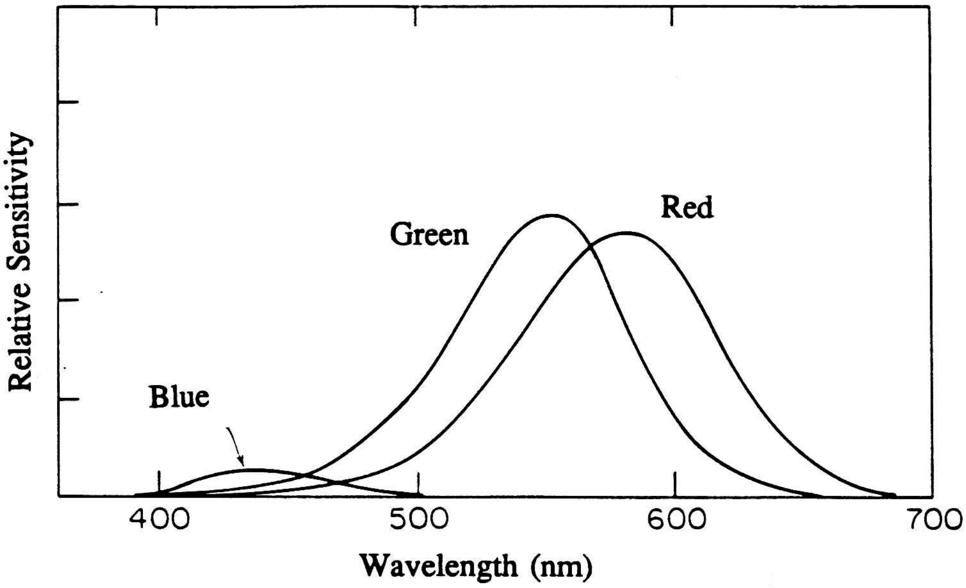


Figure 5: Typical receptor absorption curves of pigments in human retinal cones.

visible colour spectrum. Recall that a TV camera takes colour images by separately sampling three different filtered images at the same high resolution that would be required for monochrome viewing. The retina has adopted a number of strategies which allow satisfactory colour perception without significantly degrading monochrome acuity. The combined mosaic of red- and green-sensitive cones comprise 90% of the cone population and mediate high-resolution vision. The substantial overlap in the absorption spectra of the red and green sensitive cones, coupled with the relatively smooth reflectance spectra of images of real scenes, means that the outputs of the red and green cones are strongly correlated. This makes it possible for the red and green outputs to be effectively combined to produce luminance information. This luminance data is more finely sampled than the sampling any one of the populations taken individually could produce. It is degraded little by the offset in the spectral sensitivities of the red- and green-sensitive cones. Colour information at a somewhat lower resolution is still available as a difference signal between the interspersed red and green populations of cones. As well as being at lower resolution it will be somewhat noisier than the luminance information because of the smaller differences involved between the red and green sensitivities.

In contrast to the dual function of red and green cones, the blue-sensitive cones which comprise the remaining 10% of the cone population contribute little to spatial vision, but extensively to colour perception. Their absorption spectrum is strongly shifted from that of the red and green and thus provides a strong colour difference signal. Blue-sensitive cones are particularly sparse in the very centre of the fovea, which minimizes the loss in resolution that would be caused by the interruption of the red-green mosaic. It might be expected that the paucity of blue-sensitive cones in the fovea would cause aliasing of blue components of incident light. Aliasing can be demonstrated in the lab using interference techniques, but there are two reasons why it is avoided in the natural course of events. Firstly chromatic aberrations of the blue components cause strong blurring of these components without substantial deleterious effects on the spatial acuity mediated by the red/green combination. Of course this also contributes to the low spatial resolution of colour perception (maximum acuity is 6-10 cycles per degree. Nevertheless the brain seems to be able to combine the high resolution luminance information with this low resolution colour information to produce a unified high resolution spatial colour percept which is accurate in most cases. Two cases where the eye is "fooled" by high resolution colour patterns are the herringbone pattern of coloured tweed viewed at a certain distance and the "pointillistic" style of painting. Here the eye can distinguish the spatial variation of luminance, but the colours of the individual spatial features cannot be resolved and "run" or bleed into each other. The second way in which the blue mosaic is protected from inaccurate percepts due to aliasing arises from the fact that blue-sensitive cone mosaic forms a lattice which is somewhere between being perfectly regular and perfectly random. The net effect of this is that frequencies above the nominal Nyquist limits (given by the average inter-cone spacing) are not converted into conspicuous Moire patterns, but instead are scattered into broadband noise

2.4 The Periphery or Extrafovea

Outside the fovea the average spacing between cones increases rapidly. At 4° eccentricity the average sampling rate has dropped by a factor of three. By 10° the spacing is an order of magnitude down on the foveal value. Optical quality on the other hand, declines slowly, suggesting that aliasing may be an important factor in the visual performance of the extra fovea. Apart from change in average cone spacing there is a striking change in the cone mosaic with increasing eccentricity from the fovea. By $2 - 3°$ the cone mosaic has "degenerated" into an almost random lattice with no obvious regularity and with the space between the cones filled with rod receptors. The lattice is not completely random however. A perfectly random (2-dimensional Poisson) array would have no Nyquist limit in the sense in which a regular array does, and therefore it would never produce Moiré effects due to aliasing. The cost is a scattering of spectral energy of all frequencies into a “veil of white noise”. Yellott has shown, by taking the optical power spectra of cone array sampling points, that the cones provide a novel form of optimal spatial sampling. The semi-irregular cone array is optimal in the sense that the minimal noise is introduced for spatial frequencies below the nominal Nyquist limits (computed on the basis of a regular array with the same local sample spacing or sample point density). For spatial frequencies above the local Nyquist limits, conspicuous Moiré patterns are avoided by scattering the pattern energy into broadband noise.

There are conflicting views about whether the irregularity of the extrafoveal cone mosaic is a device which has been selected for by evolution to defeat aliasing distortion, or if it is simply a residual disorder as a result of a lack of selection pressure for further regularity. Whichever is nearer to the truth there are a number of reasons why aliased patterns (for a regular array) or aliasing noise (for an irregular array) may not be a significant problem for extrafoveal vision. Most of the time the eye is not well accommodated to objects projected on the periphery: to scrutinize something, the eye makes a *saccade* to allow the image to be projected onto the fovea and then accommodates to bring the foveal image into focus. It is also suspected that light scatter in pre-receptoral layers of the peripheral retina causes a blurring of the projected image in these regions. Finally it has been shown that natural scenes have most of their power appearing at low frequencies. These three factors all have the effect of giving low contrast at higher spatial frequencies and consequently reducing problems due to aliasing noise.

2.5 Extrafoveal function

A more important factor in the development of the extrafoveal cone mosaic may be the function of the periphery. The primary function of the peripheral retina in daylight is the detection of objects particularly the detection of any type of motion. An eye/head fixation movement then allows the fovea to scrutinize the region where something was detected. This function requires the periphery to maximize contrast at sub-Nyquist frequencies, even at the expense of aliasing noise. Good contrast means that the extrafoveal cones must have good quantum efficiency and therefore need to be larger than foveal cones. Even though aliasing could be avoided by reducing the extrafoveal cone size to that of foveal cones and then reducing the sample spacing, this biologically possible step is not taken. It seems that in the case of the extrafoveal retina, a premium is placed on detection, with several factors conspiring to reduce the cost of this in terms of aliasing noise. In fact, the close match between optical quality and sample spacing in the fovea may be the exception rather than the rule. There is widespread undersampling of the cone mosaic of many other species of animals which do not have the specialized scrutinizing neural (*parvo*) subsystem of primates. The decrease in sample spacing of the foveal mosaic to the limit allowed by the refractive indices of available biochemical materials may have been closely associated with the development of the parvo system in primates. The much older (in evolutionary terms) *magno* neural subsystem seems to be responsible in primates for the detection and processing of motion, stereo and depth cues and for gestalt grouping effects. It is not capable of the detailed and sustained scrutiny of patterns or objects necessary for fine manipulation. The newer, and in the primate, much more extensive parvo system is capable of sustained and detailed pattern scrutiny (though this ability is substantially degraded by motion and lacks the gestalt abilities characteristic of the magno system). It is impossible to say at this stage if any one of the three adaptations of (i) fine cone sampling for high acuity (ii) the parvo system for object/pattern scrutiny or (iii) the development of dexterous manipulation abilities, provided the impetus for the development of the others. It does nevertheless seem that they are closely related in some manner. In contrast to the fovea where post-receptor neural processing hardly effects acuity at all, neural mechanisms in the peripheral retina impose substantial limitations on visual acuity. Extrafoveal acuity falls well below the cone mosaic's average Nyquist limit and this is reflected in the 5-to-1 ratio of cones to ganglion cells in the far periphery. This decrease in acuity caused by the subsequent neural processing may not be surprising in view of the acknowledged role of the periphery in detection rather than examination, and the fact that the processing in the fovea demonstrates that acuity can be maintained, if required. It also further reinforces the rejection of the notion of a retina which simply transduces and codes "images". Clearly the retina carries out the role of detecting appropriate information available in the optic array. Our interest is in knowing what the "appropriate information" is, and how it can be extracted or made explicit. The neural processing of the retina and how it develops may give clues which would help to answer these questions.

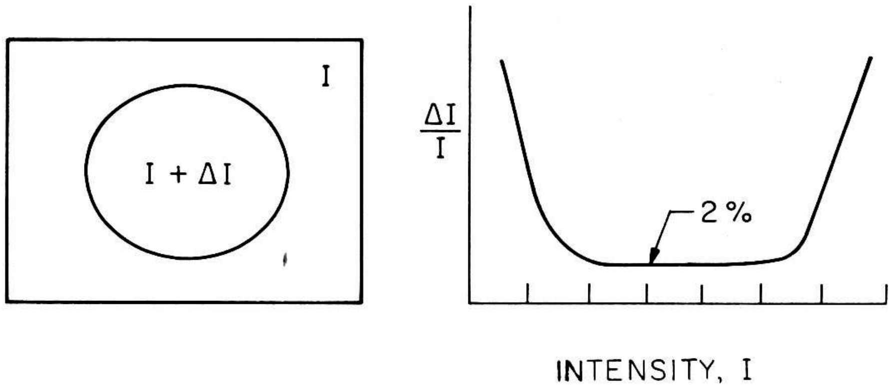
3 Some Psychophysical Properties of the Human Visual System

3.1 Contrast Sensitivity

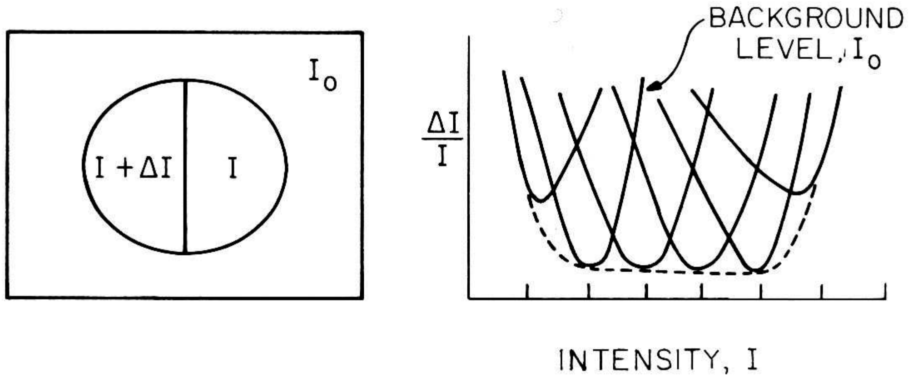
Most natural objects are visible because they absorb and reflect constant proportions of the light falling on them in various parts of the visible spectrum. This is a physical property determined by the molecular makeup and surface characteristics of the matter comprising the object. It is a property which is invariant with normal changes in ambient illumination of natural scenes and is encoded by contrast or relative intensity (possibly as a function of frequency). Objects “look” the same to a contrast encoding system as the mean intensity changes over a wide range of values. The ability of individual photoreceptors to adapt their sensitivity to match the ambient light level is important for coping with the 10,000-fold range of light intensities during the day. This type of adaptation automatically allows them to encode contrast in their outputs rather than absolute intensity. The ratio of just noticeable intensity difference to local mean intensity, called the Weber fraction $\Delta I/I$, is found to have a nearly constant value of 2% over a wide range of intensities in human vision. As this ratio is equivalent to a small change in the logarithm of intensity, $\Delta(\log I)$, then just perceivable contrasts at any mean intensity are directly related to changes in the log of intensity. Consider a response where inputs are considered equivalent if the quantity described by contrast, when superimposed on a background signal that increases with mean intensity, is constant. This type of response is equivalent to a logarithmic transformation of input intensity. It is a transformation which is found in a number of different types of insect photoreceptors and vertebrate cones. Other models of the contrast function of the human visual system include the root function $I^{1/n}$, with, for example, $n = 3$.

3.2 Mach Bands

In addition to the compressive non-linear point transformation of intensity which is modelled as a logarithmic or root function, there are spatial and temporal interactions involved in the transduction of physical luminance to perceptual brightness. Where the underlying physiology is understood very well, for example in the receptors on the compound eye of certain types of fly, these interactions have been explained in terms of adaptive noise reduction and predictive coding. We briefly outline some of the results of these interactions here, without justifying the particular reasons in the more complex case of the human eye.



(a) NO BACKGROUND



(b) WITH BACKGROUND

Figure 6: Schematic showing the range over which the Weber relation holds. Any of the curves for fixed I_0 in (b) is comparable with the dynamic range of most electronic imaging systems.

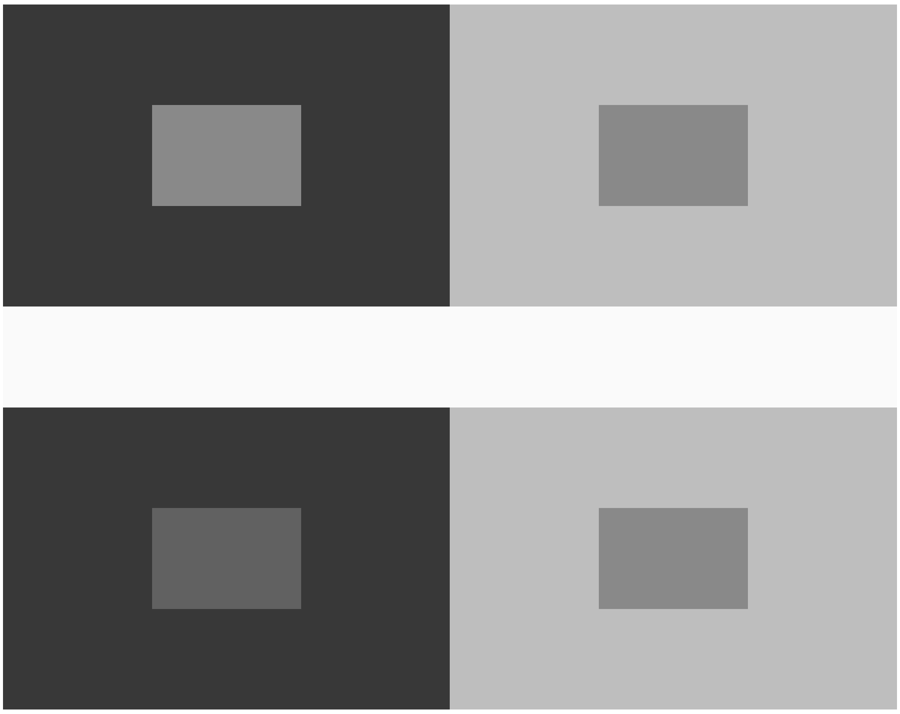


Figure 7: In the top pair of images, the small center squares have equal luminances, but do not appear equally bright. In the bottom pair of images, the center squares have different luminances but appear equally bright.

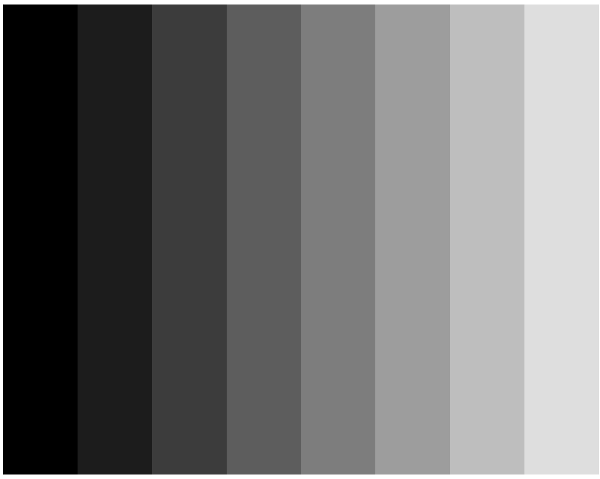


Figure 8: The luminance within each of the vertical bars is constant. However, the perceived brightness is not.

In the *Mach band* diagram shown below, the subjective brightness across the vertical grey bars is not uniform in spite of the fact that the individual bars have constant luminance. This phenomenon is known as the *Mach band illusion*. See if you can explain what transformation would give the spatial variation of brightness observer.

The mechanism which gives rise to the Mach-band effect is referred to as *lateral inhibition*. What would the spatial impulse response for a "laterally inhibited" spatial interaction look like?

3.3 The Modulation Transfer Function

The "lateral inhibition" interaction also effects how different spatial frequencies are perceived. The net result is that we need a very high contrast to see very low and very high spatial frequencies. On the other hand, we are sensitive to small contrast changes if they appear in mid-range spatial freuqencies. The overall effect is illustrated by the diagram below which shows a sine-shape grating which varies in frequency across the image and decreases in contrast down the image. The threshold of just-noticable contrast as a function of spatial frequency is referred to as the *modulation transfer function*. The peak sensitivity is usually in the range 3-10 cycles per degree.

The modulation transfer function for colour perception in the human visual system is shifted down in spatial frequency from that for sensitivity to intensity changes. This is the principal reason why it is possible to sub-

Figure 9: The luminance within each of the vertical bars is not constant. See if you can determine how luminance varies as you traverse the image from left to right.

sample the U and V components of a YUV frame representation in image and video coding schemes without affecting the perceived quality.

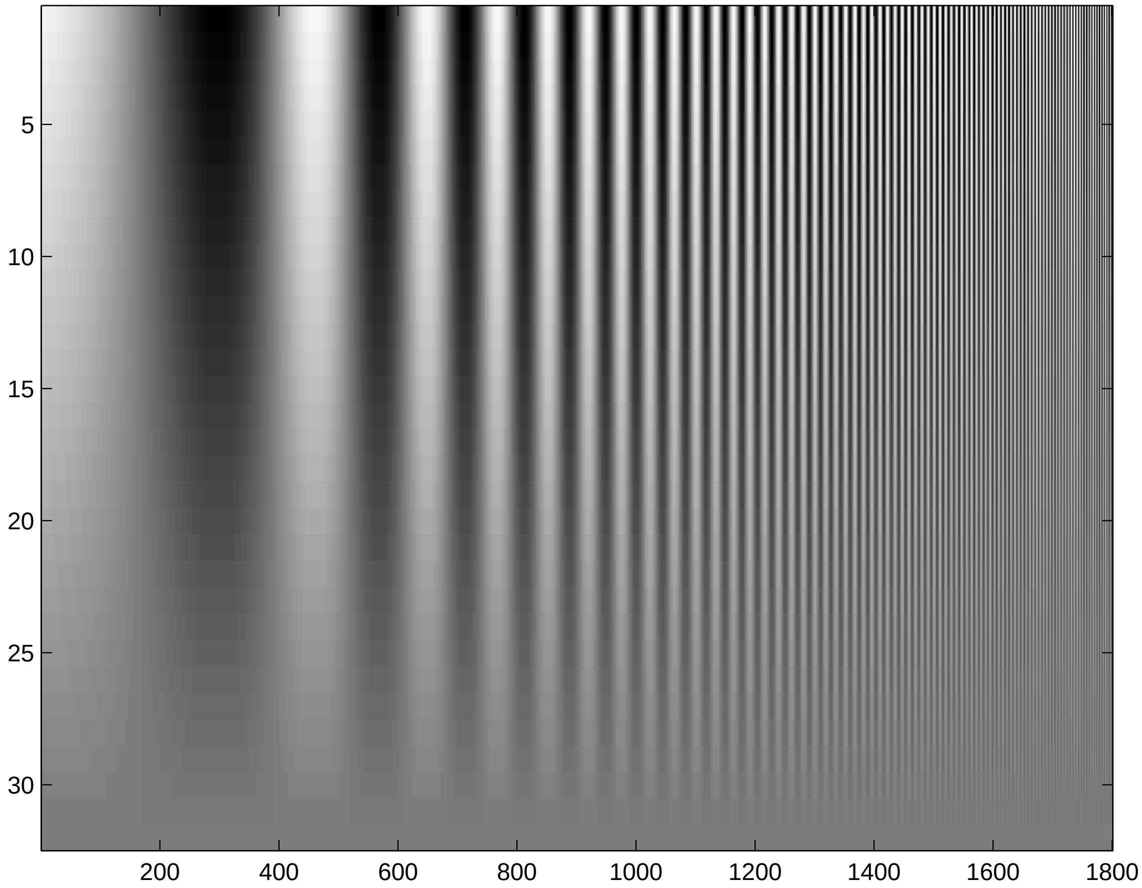


Figure 10: Image a sine-shape grating which varies in frequency across the image and decreases in contrast down the image. Try to trace the curve of just-noticable luminance change. Ignore the Matlab axis labelling.

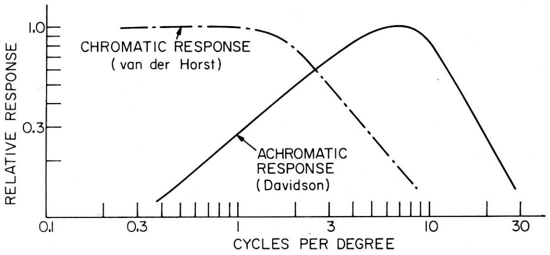


Figure 11: Plot of the modulation transfer functions for colour and luminance contrast sensitivity of the human visual system.