Perception, 1986, volume 15, pages 17-25

Digitized images: what type of grey scale should one use?

Fred Kingdom, Bernard Moulden
Department of Psychology, University of Reading, Whiteknights, Reading RG6 2AL, UK
Received 9 October 1985, in revised form 22 January 1986

Abstract. One of the issues faced by engineers when designing a system which records an external event and represents it in the form of a digitized image on a VDU screen is which type of grey scale to use. An experiment is described which compares, in a simulated digitized image, the effect of a linear and a logarithmic grey scale on the detectability of a straight-line signal embedded in visual noise. It was found that both bright and dark signals were detected more easily with the linear scale. A signal detection theory analysis was carried out to compare human performance with that of an 'ideal' observer who performed the detection task with a filter spatially matched to the signal. It was found that the model of performance for this ideal observer accounted well for the results provided the assumption of a linear transformation of luminance was made. The analysis showed that the superiority of the linear over the logarithmic grey scale was simply due to the higher signal-to-noise ratio of the signals in the former.

1 Introduction

In the past two decades the rapid development of digital image processing techniques has provided both a tool for visual scientists and an opportunity for the practical application of the results of their investigations. In digital image processing external events such as visual scenes, underwater sounds, and body scans are processed by computer and displayed as images on a visual display unit (VDU). Typically an operator is required to detect, recognize, and analyze target shapes within those images. Clearly the design of such a system will be partly guided by the need to maximize the subjective quality of the images in general and the visibility of targets in particular. There is now a large body of literature in which the particular merits of various types of image processing or image enhancement, obtained either by subjective impressions of image quality (Todd-Pokropek and Huggett 1981; Chao et al 1983; Peli 1984) or by specific detection and recognition tasks (Chambers and Courtney-Pratt 1969, Judy and Swensson 1981; Chao et al 1983) are assessed.

Complex displays may employ colour, or may vary as a function of time to generate flicker or movement; but typically the output image is a function of only three parameters, two being the $X-Y$ spatial coordinates and the third being the intensity of the image at a particular $X-Y$ location. The purpose of the image is to provide a symbolic visual representation of some external state of affairs, and the broad issues facing the designer may be simply stated: how may these three dimensions of variation in the image best be used to convey the required information about the external world? This fundamental question itself may be considered as having two levels. The first concerns the mapping of external information onto the three available dimensions of variation in the image: what are these dimensions to represent? In some cases the question hardly arises because the mapping is so obvious, as in the case of radiography: in an x-ray image the two spatial dimensions are used to represent directly the spatial arrangement of the source object, and intensity is used to represent opacity. In other cases, however, there is no such obvious source-object-to-image compatibility; one spatial dimension of the image may be used to represent time, the other to represent frequency, and the dimension of intensity to represent amplitude, as in the familiar sound spectrograph.

Once this first question has been resolved the second question becomes clear: what should be the *nature* of the mapping of object properties onto image dimensions? How should different values in the object dimensions (time, frequency, amplitude, and so on) be represented by different values in the image dimensions? For example, should equal spatial intervals in the image represent equal intervals in object frequency (a linear scale) or equal ratios in object frequency (a logarithmic scale)? It is to one aspect of this second class of issues that the present study is directed. When variations in intensity in the image are used to represent variations in the magnitude of an object dimension, what is the best type of grey scale to use?

When an external event—essentially an analog stream of, for example, intensity across two-dimensional space for a visual event or amplitude across frequency-time for an auditory one—is digitized, it is divided into 'bins', each bin representing a specified amplitude range in the input. This process is often referred to as 'quantization'. After quantization each bin is allocated a single luminance in the output image, with the total number of luminances, or grey levels, corresponding to the total number of bins. The grey scale may define the absolute values of those allocated luminances, or simply, and perhaps more importantly, the function relating grey-scale values to luminance.

Our aim was to investigate the effect of different types of grey level on signal detectability, first as a matter of straightforward empirical interest and second in an attempt to derive a simple theoretical rule to account for any difference in performance. We chose to compare two grey-level scales, one in which the intensity difference between successive grey-scale values was a constant (the linear scale) and one in which the intensity difference between successive grey levels increased by a constant proportion. Because such a scale can be described as consisting of equal logarithmic intervals we refer to it as a logarithmic scale. This scale was selected because it has commonly been chosen by display engineers (eg Chanda et al 1984).

In view of some recent experimental work (Legge and Kersten 1983) suggesting that decremental signals are in some circumstances more detectable than incremental signals, we also investigated the effect of signals which were on average either brighter or darker than the background against which they were presented.

2 Method

2.1 *Subjects*

Three male undergraduates participated. All had normal vision and were naive as to the purpose of the experiment.

2.2 *General characteristics of the displays*

Instead of taking an input trace and then quantizing it, we simulated the effect of this by means of an algorithm which directly computed the output display. The precise way in which this was done is discussed below.

The stimuli were generated by a Supervisor 214 (Gresham Lion PPL) graphics/image system interfaced to a PDP 11/40 host computer and displayed on a Digivision TV monitor. The luminances of the 256 grey levels available in the system were measured with a United Detector Technology microphotometer (model 81) focused onto the phosphor of the monitor when it displayed a large homogenous patch of pixels. Each pixel was approximately circular with a diameter of 1.27 min visual arc.

2.3 *Grey-level distribution of signal and noise*

The displays all contained eight grey levels, the actual luminances of which were selected from the 256 available according to the type of grey scale under investigation.

Each pixel was randomly allocated one of the grey levels, i, with a probability P_i determined by the binomial rule where

$$P_i = \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r}, \qquad i = r+1; \quad N = 7. \tag{1}$$

The underlying shape of the distribution of grey-level probabilities is determined by the parameter p in equation (1), where p can range from 0 to 1. When $p = 0.5$ the distribution is symmetrical; as p increases from 0.5 to 1 the distribution becomes increasingly skewed to the left. The noise was generated with p set to 0.5 and figure 1a shows the resulting histogram of grey-level probabilities. Five signals of different strengths were generated with p values of 0.56, 0.57, 0.58, 0.59, and 0.6. When the luminances were allocated on an ascending scale (1 - 8) the average intensity of the signals was greater than that of the noise owing to the larger proportion of pixels with high luminances, as can be seen from the grey-level histogram for a bright signal with $p = 0.58$ shown in figure 1b. When a descending scale was used (grey levels 8 - 1) the signals were darker owing to the larger proportion of pixels with lower luminances. The signal p values were chosen to produce a range of performance giving approximately between 10% and 90% correct signal detections as determined in pilot experiments.

The average luminance $\bar{L}$ and variance σ^2 of the signals and the noise were calculated numerically with the following equations:

$$\bar{L} = \sum_i P_i L_i, \qquad \sigma^2 = \sum_i P_i (L_i - \bar{L})^2,$$

where P_i and L_i are the probability of occurrence and luminance, respectively, of the ith grey level.

One consequence of generating stimuli in this way is to make the distribution of luminances in the signals asymmetrical: either higher luminances are more common than lower luminances (signals brighter than the background) or the reverse (signals darker than the background). Previous investigations have used only symmetrical distributions (eg Sullivan 1983) and the present paper may be seen, in part, as extending their findings to asymmetrical distributions.

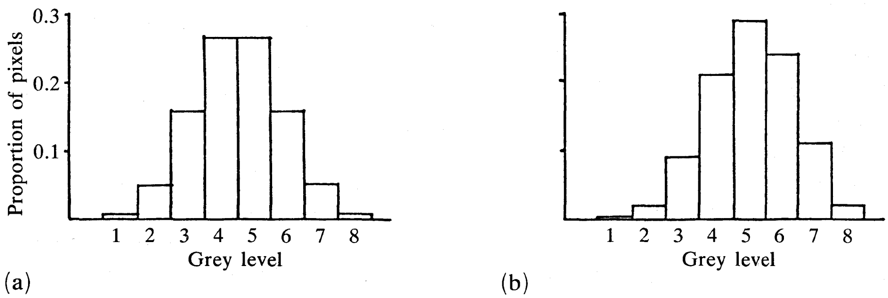

Figure 1. The grey-level distribution for (a) the noise background and (b) a signal with $p = 0.58$. The values on the abscissa do not represent actual luminances but are nominal labels for the eight grey levels. The actual luminance values depended on the type of grey scale employed.

2.4 *Spatial characteristics of signals and noise*

The noise background was a slab of 512 pixels wide × 256 pixels high, subtending 10.2 deg × 5.1 deg, and was randomly selected from six pages of 512 × 512 pixels, freshly generated before each experimental session and stored in the memory planes of the Supervisor 214. The signal was a vertical line 1 pixel × 256 pixels which replaced one of the 512 noise 'lines'. On each trial it was freshly generated.

2.5 *Grey scales*

By allocating values from the palette of 256 to the eight digital values it was possible to produce a grey scale of any desired specification. The two scales of interest here were (i) a linear scale (equally spaced luminances) and (ii) a logarithmic scale (equal ratios of luminances). Since a linear scale with the same luminance *range* as that of a log scale results in a mean luminance approximately 1 log unit greater in our images (see below), it was decided to use a third scale, again linear but with the same mean luminance as that of the log scale. This served to check that any difference in performance found for the two scales was not attributable to the substantial difference in overall mean luminance.

The logarithmic scale consisted of eight luminances in the range 0.1 – 62 cd m^{-2}, resulting in a noise background with a calculated mean luminance of 3.76 cd m^{-2}. The first linear scale had the same range but a mean luminance of 29.1 cd m^{-2}; the second linear scale had the same mean luminance as that of the log scale but a range 0.1 – 7.6 cd m^{-2}. For each of the three types of scale used there were two conditions, one in which the scale was allocated in an ascending direction, ie mapped onto digital values 1 – 8, the other in which the scale was allocated in a descending direction, ie 8 – 1. Because the grey-level probability distribution for the noise background was symmetrical (figure 1a) it made no difference whether a given scale was allocated in a descending or an ascending way. On the other hand, for the signals an ascending scale resulted in a signal brighter than its background and a descending scale in one that was darker. Thus there were six main conditions in all, and five signal strengths for each condition.

2.6 *Procedure*

After familiarizing the subject with the experiment through sufficient practice trials the subject was seated 120 cm from the monitor. In the interval between each trial, in which the subject viewed the screen with a homogenous field of mean luminance the same as that of the noise, a signal replaced one of the randomly selected 510 horizontal locations of the slab of noise that had itself been randomly selected from the six available pages in the Supervisor 214. The two vertical noise-line positions on the extreme boundaries were not considered as potential signal locations, thus ensuring that the signal was certain to be surrounded by at least one noise line on each side. During the trial the subject was required to find the signal, which he was told was the line which had overall the greatest (or least) intensity. The subject indicated the horizontal location of the signal with a box cursor which could be positioned precisely over a noise or signal line by means of a tracker ball which he controlled. When he was satisfied with his choice he pressed a button and his response was recorded. At this point the stimulus disappeared and was replaced by another 4 s later.

For each of the six conditions each subject performed three hundred trials divided into six sessions of fifty trials each. During a single experimental session the five different signal intensities (ten trials of each) were randomly presented but there was only one type of grey scale per session. The order of presentation of conditions was randomized.

3 Results

Figure 2 shows the means of the percentage correct scores for each condition and signal strength, each data point representing the average between the three subjects, ie from 180 trials. As can be seen from the figure, performance with the two linear grey scales, for both bright and dark signals, is superior to that of the logarithmic scale. A four-way analysis of variance (ANOVA) with grey scale, type of signal intensity (bright versus dark), signal strength (p value), and subjects as factors showed that there was a main effect of type of grey scale ($F_{2,4} = 43.3$, $p < 0.01$). A separate four-way ANOVA in which the results from the logarithmic scale were omitted showed no significant main effect of grey scale ($F_{1,2} = 0.36$, $p = 0.6$). As can be seen from the figure, the results from the two linear scales fall very closely together for both bright and dark signals.

As regards the difference in performance between bright and dark signals, the first four-way ANOVA showed no significant overall difference in performance ($F_{1,2} = 0.3$, $p = 0.63$), but a significant interaction between type of signal intensity and type of grey scale ($F_{2,4} = 31.52$, $p < 0.01$). As can be seen from figure 2, performance with dark signals was slightly better than with bright signals for the linear scale condition but considerably worse for the log scale. If one analyzes the results for the two types of scale separately, the superiority of the dark signals with the linear scale is found not to be significant ($F_{1,2} = 2.05$, $p = 0.29$). On the other hand, the superiority of the bright signals for the log scale data is significant ($F_{1,2} = 50.9$, $p < 0.05$).

Although from a practical point of view the performance of our subjects as measured by the percentage of correct signal detections is sufficient to guide any engineer designing a system producing images similar to ours, theoretical considerations demand a deeper analysis of the results. We have therefore used a signal detection theory (SDT) analysis (Green and Swets 1966) to convert our percentage correct values to a measure of sensitivity for the subjects, d'_e, for each of the various conditions of the experiment. A number of studies have been carried out with SDT analysis for the detection of luminance increment against visual noise (eg Cohn and Lasley 1975; Swets and Birdsall 1978; Burgess et al 1982; Sullivan 1983). However, these studies have invariably used the paradigm in which a completely homogenous signal is added to gaussian white noise ie noise with a very large number of grey levels, in order to satisfy the requirement that the signals and noise be gaussian with equal variance, the simplest case to deal with mathematically. Although our study used displays with just eight grey levels and the noise and signals differed in the shape of their underlying luminance distributions, it is shown below that a conventional SDT analysis can be made.

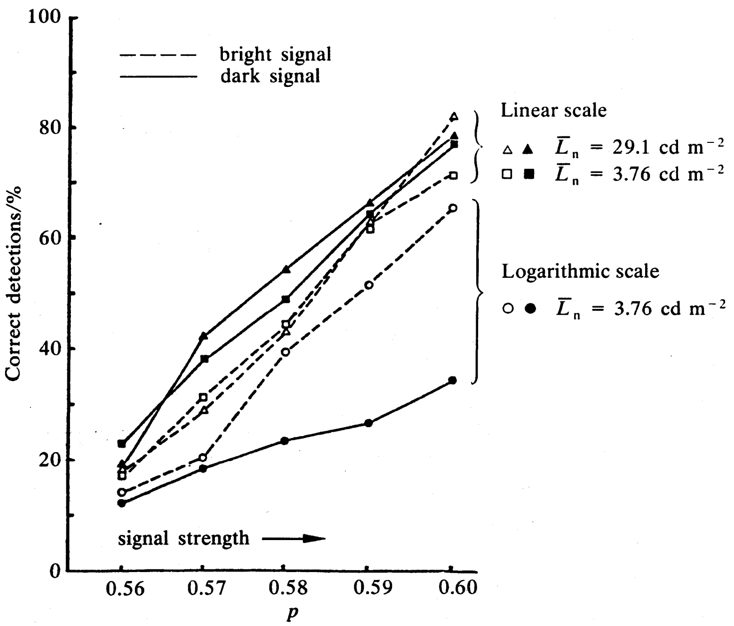

Figure 2. The percentage of correct signal detections as a function of signal intensity, type of grey scale, and bright versus dark signal. Signal intensity is plotted on the abscissa and is here measured simply by the parameter p which indicates the degree of skewness of the binomial distribution of grey levels in the signal.

The SDT analysis allows us to compare human performance with that of an 'ideal' observer who performs the detection task using a filter whose spatial dimensions exactly match those of the signal. The output of the filter is proportional to the mean luminance of all the pixels which fall within it and the assumption is that the filter is used as a travelling window to measure the average intensity of all the potential signal candidates in the display and to indicate as the signal the candidate with the highest value. The ideal observer used here is less general than the one which is defined as utilizing *all* the statistical information available in the display for signal detection (Barlow and Reeves, 1979; Burgess et al 1982), since it does not utilize the known differences in the underlying *shapes* of the grey-level distributions of the signal and noise. We have shown in a more recent experiment (to be published) that with otherwise identical displays extremely large differences in the shape of the underlying distributions are required before detection by human observers can take place in the absence of overall luminance differences. The ideal observer used here, therefore, is best thought of as a proposed model of performance for the hypothetical situation in which no noise is present other than that due to the statistical variation in the output of the matched filter that is employed.

The performance of this ideal observer is measured by d_{i}' which is defined as:

$$d_{\mathrm{i}}' = \frac{\overline{L}_{\mathrm{s}} - \overline{L}_{\mathrm{n}}}{\sigma_{\mathrm{n}} A^{-1/2}} ,$$

where $\overline{L}_{\mathrm{s}}$ and $\overline{L}_{\mathrm{n}}$ are the mean luminance of the signal and noise respectively, σ_{n} is the standard deviation of noise pixel luminance, and A is number of pixels making up the signal, in this case 256. The term in the denominator is essentially the standard deviation of the mean luminance of a line of 256 noise pixels. In other words, these two factors (i) the *difference* in intensity between the signal and noise and (ii) the *variance* in intensity of the output of the matched filter to the noise lines, are assumed in this model of detectability to be the two factors affecting performance. The term d_{i}' is also the signal-to-noise ratio (SNR) of the output of the matched filter expressed in standard deviation units of the noise distribution.

To convert the percentage correct measures into d_{e}' values, we used an iterative search procedure to obtain d_{e}' by numerical evaluation of the following formula, taken from Green and Swets (1966), by means of which one can obtain $\mathscr{P}$ (proportion of correct signal detections) from d_{e}': [(1)]

$$\mathscr{P} = \int_{-\infty}^{+\infty} \phi(x - d_{\mathrm{e}}')\, \Phi(x)^{M-1} \mathrm{d}x ,$$

where ϕ and Φ represent the ordinate and area under the lower tail of the unit normal distribution and M is the number of signal alternatives, in this instance 510. The assumption is that the distributions of the signal and noise are normal with equal variance. Although the grey-level distributions in the signal and noise have a binomial distribution, the sample mean luminances of the noise and signal lines, each containing 256 pixels, will be normally distributed. We have shown empirically (to be published) with yes/no ROC (receiver operating characteristic) curves that the ratio of standard deviations of the signal and noise are sufficiently close to unity for the equal variance assumption to be made.

To compare human performance at this task with that of our ideal observer we have used a measure of efficiency, F, where $F = (d_{\mathrm{e}}'/d_{\mathrm{i}}')^2$. This measure has been used for the measurement of the efficiency with which humans process luminance increments

(1) We would like to thank Professor Curnow of the Department of Applied Statistics, Reading University, for help with the necessary software.

(Sullivan 1983; Burgess and Ghandeharian 1984), sine-wave intensity-modulated signals (Burgess et al 1982), dot density (Barlow 1978), and symmetry in random-dot patterns (Barlow and Reeves 1979).

The first reason for measuring efficiency for the various conditions of this experiment is that it allows us, for each grey scale, to disambiguate the differences in performance which could be attributable to differences in the SNR from those attributable to differences in visual processing. If all the efficiency values collapse onto a single point it indicates that the observed difference in performance in terms of percentage correct detections was simply a consequence of the different SNR associated with each of the different grey scales. If all the efficiency values are not identical, then this would suggest that there is something fundamentally different about the way in which the visual system processes displays with different grey scales.

The various measures of F are shown in figures 3a and 3b: figure 3a for the bright signals and figure 3b for the dark signals. As can be seen, the F values in figure 3a fall very closely together around a mean value of 18%. The F values for the dark signals are not as close together and have an average value of 22%. The log scale F values for the dark signals are noticeably greater than their linear scale counterparts, although it is not clear why this should be so.

A four-way ANOVA on the overall F measures shows no significant differences between type of grey scale ($F_{2,4} = 1.4$, $p = 0.34$), or signal strength ($F_{4,8} = 1.16$, $p = 0.39$). The difference between bright and dark signals is just nonsignificant ($F_{1,2} = 16.3$, $p = 0.06$).

Linear regression analyses with the method of least squares were carried out to ascertain to what extent the d'_i values could account for the variance in performance between the various grey scales. The analyses were carried out with the mean values of d'_e across subjects and separately for both bright and dark signals. The resulting coefficients of determination, R^2 (the proportion of variance in performance across conditions that is attributable to d'_i), are as follows: for bright signals $R^2 = 0.96$; for dark signals $R^2 = 0.97$. The model of signal detection appears to account for the data very well.

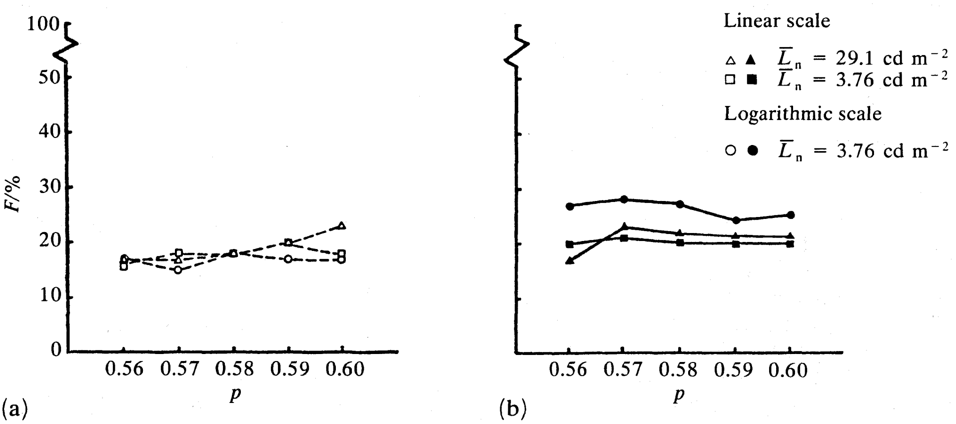

Figure 3. The measures of efficiency, F, with which the performance of the human subjects is compared with that of an 'ideal' observer using a filter whose spatial dimensions match the signal exactly: (a) bright signals, (b) dark signals.

4 Discussion

4.1 *Type of grey scale*

The percentage correct results provide good evidence that for practical displays of the kind employed in this experiment a linear scale is preferable to a logarithmic one.

Performance with our subjects was on average 8% better with the linear scale for bright signals and 28% better for dark signals.

The difference between the linear and logarithmic grey scales is attributable to the greater signal-to-noise ratio of the linear scale signals as shown by the respective d'_i values; this is discussed in more detail below in the context of a model for these data.

4.2 *Bright versus dark signals*

Some previous findings have shown that luminance decrements are more visible than increments when simple stimuli such as bars or gratings are used (Cohn and Lasley 1975; Legge and Kersten 1983). Some of our data show a similar slight trend in this direction but it is nonsignificant. The phenomenon found with simple stimuli is held to be a consequence of a logarithmic transform of input intensity at an early stage in visual processing (Legge and Kersten 1983). The model of detectability proposed here assumes a linear transform of input intensity and therefore would not predict any difference between bright and dark signals. It is not clear why the two different kinds of experiment appear to give different results.

4.3 *Model of signal detection*

Our data are best described by a model which does not require a nonlinear transform at the input stage, as the results of some studies seem to suggest (Maudarbocus and Ruddock 1973; Legge and Kersten 1983; Morgan et al 1984). Rather, it seems that, at least under the conditions of our experiment, the system can best be described in terms of a linear input stage followed by a linear summation of luminances falling within the integration region of an elongated receptive field.

The simple model of signal detection proposed here, namely that detection depends on a filter whose spatial dimensions match those of the signal and whose output depends solely on the total amount of light energy falling within its integration field, provides a good description of the data. Detectability depends upon two factors; the first is the relative size of the output of the filter at the location of the signal and the second is the variability of its output at all other locations. The filter can be thought of as an elongated receptive field which integrates the luminance of points falling within its excitatory region. Our results, of course, do not tell us anything about the relationship between the actual spatial dimensions of the signal and its detectability or to what extent such a relationship can be explained in terms of known properties of visual receptive fields. We are currently carrying out experiments on this.

Acknowledgement. This research was funded by the Procurement Executive, Ministry of Defence.

References

Barlow H B, 1978 "The efficiency of detecting changes of density in random dot patterns" *Vision Research* **18** 637–650

Barlow H B, Reeves B C, 1979 "The versatility and absolute efficiency of detecting mirror symmetry in random dot displays" *Vision Research* **19** 783–793

Burgess A E, Ghandeharian H, 1984 "Visual signal detection. II. Signal-location identification" *Journal of the Optical Society of America* **1** 906–910

Burgess A E, Jennings R J, Wagner R F, 1982 "Statistical efficiency: A measure of human visual signal-detection performance" *Journal of Applied Photographic Engineering* **8** 76–78

Chambers R P, Courtney-Pratt J S, 1969 "Experiments on the detection of visual signals in noise using computer generated signals" *Photographic Science and Engineering* **13** 286–298

Chanda B, Chaudhuri B B, Dutta Majumder D, 1984 "Some algorithms for image enhancement incorporating human visual response" *Pattern Recognition* **17** 423–428

Chao B P, Beaton R J, Snyder H L, 1983 "Human performance evaluations of digital image quality" in *Proceedings SPIE International Society of Optical Engineering (USA): Advances in Display Technology III* **386** 20–24

Cohn T E, Lasley D J, 1975 "Spatial summation of foveal increments and decrements" *Vision Research* **15** 389–399

Green D M, Swets J A, 1966 *Signal Detection and Psychophysics* (New York: John Wiley)

Judy P F, Swensson R G, 1981 "Lesion detection and signal-to-noise ratio in CT images" *Medical Physics* **8** 13 - 23

Legge G E, Kersten D, 1983 "Light and dark bars: Contrast discrimination" *Vision Research* **23** 473 - 483

Maudarbocus A Y, Ruddock K H, 1973 "Non-linearity of visual signals in relation to shape sensitive adaptation responses" *Vision Research* **13** 1713 - 1738

Morgan M J, Mather G, Moulden B, Watt R J, 1984 "Intensity-response nonlinearities and the theory of edge localization" *Vision Research* **24** 713 - 719

Peli E, 1984 "Image enhancement for the visually impaired" *Optical Engineering* **23** 47 - 51

Sullivan G D, 1983, "Perceptual filters" in *Physical and Biological Processing of Images* eds O J Braddick, A C Sleigh (Berlin: Springer) pp 115 - 126

Swets J A, Birdsall T G, 1978 "Repeated observation of an uncertain signal" *Perception & Psychophysics* **23** 269 - 274

Todd-Pokropek A, Huggett S, 1981 "Visual perception and the assessment of the quality of medical images" in *2nd International Conference on Visual Psychophysics and Medical Imaging, Brussels, Belgium, 2-3 July* (New York: IEEE) pp 95 - 101