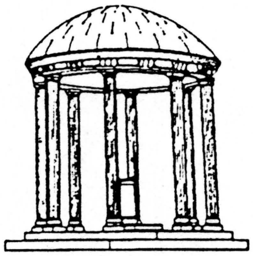
Standardization of Electronic Display Devices Based on Human Perception

December, 1987

Technical Report 88-002

Robert C. Cromartie, R. Eugene Johnston,
Stephen M. Pizer, and Diane Rogers

The University of North Carolina at Chapel Hill
Department of Computer Science
Sitterson Hall, 083A
Chapel Hill, NC 27599-3175



Submitted to Human Factors, 1987

Standardization of Electronic Display Devices Based on Human Perception

ROBERT C. CROMARTIE, Dept. of Computer Science, R.EUGENE JOHNSTON,[1] Dept. of Radiology, STEPHEN M. PIZER, Depts. of Computer Science and Radiology, and DIANE C. ROGERS, Dept of Radiology, University of North Carolina, Chapel Hill, N.C.

A method is presented for standardizing devices used to display images. It incorporates knowledge of human visual perception and display technologies to produce perceptually linear displays for which equal changes in an image intensity produce equally discriminable brightness changes. This is done by determining a function from an image intensity to the driving intensity of the device. This approach provides a standard baseline for image display from which display devices can be compared, and it facilitates use of contrast enhancement routines to maximize image information for particular objectives. The experimental method used to construct the mapping function is described as well as the computer programs that produce the mapping.

1) Requests for reprints should be sent to R.E.Johnston, Imaging Division Department of Radiology, Univ of North Carolina, Chapel Hill, NC 27514.

Running Title: Display Standardization

Key words: Display standardization, Medical image display, Image perception.

INTRODUCTION

The display of digitally stored images by electronic devices is a relatively new field that is rapidly expanding for several reasons. Storing images in computers allows rapid access to, and fast transfer of images across great distances. It also provides compact representation and reliable long-term storage of images. Mathematical techniques to enhance and analyze images are available and digital displays allow the user to control many of the parameters of the display process which are predetermined for hard copy (paper or film) display.

Digital images. An image is stored in a computer as an array of numbers, each number representing the recorded intensity of one small area of the image, called a picture element, or pixel. The range of these numbers may vary considerably, depending on the imaging technique and the storage capabilities of the computer. In the course of display, the stored image is sequentially represented in the manner schematized in Figure 1. First, it must begin with an assignment of numeric stored information to numbers that represent intensities to be displayed.

---

insert fig 1 about here

The transformed image numbers, now referred to as scale indicator intensities, are mapped by a lookup table (LUT) into the <u>physical</u> <u>driving</u> <u>intensities</u>, i.e., digital driving levels, of the display device. A digital-to-analog converter (DAC) uses this as input and produces the video signals, which in turn are converted by the display device into <u>displayed</u> <u>intensities</u>, i.e., luminance. The term greyscale refers to the path traversed from minimum to maximum displayed intensity (photometric luminance) as a function of digital driving level. As the final transformation, these displayed intensities are processed by the human visual system and produce the psychological experience of brightness or perceived intensity.

When a digitally stored image is displayed on different devices, the resultant images will differ physically from each other as a function of the devices' operating characteristics. For example, with video display systems, different phosphors and video electronics may produce considerable differences in the displayed image. We are concerned that these physical differences between images may influence performance of observer tasks.

<u>Standardization.</u> We believe that an important aspect of video display use involves standardizing the displays so that perceptually, these would-be variations of images on different display systems are controlled. That is, a user's perception of a given (digitally stored) image should be equivalent across displays, despite possible differences in the displays' operating characteristics. In this paper, we describe an approach for

"perceptual standardization" of displayed images.

The most straight forward standardization procedure would result in all display systems having physically identical luminance profiles when a given stored image is displayed. This may not always be the best approach, however. For instance, when displays are viewed under substantially different ambient light conditions, the light sensitivity of the visual system changes and the luminance intensity range for the display should be adjusted accordingly. Specifically, the visual system can discriminate changes in luminance over a range of about 10 orders of magnitude, but at any time, sensitivity operates over a luminance range restricted to about 2 orders of magnitude. At higher intensities, visual performance for tasks involving spatial and temporal resolution is higher. This suggests a second reason to not require physical matching of displays as a basis of standardization. Namely, when display devices have different photometric intensity ranges, each device should extend its range as far as possible. As a consequence of these display adjustments for the characteristics of the human visual system, of course, the images displayed on different devices will not be physically identical. The goal of our proposed standardization methodology is to achieve perceptual, not physical, equivalence between devices and perceptual consistency, on a day to day basis, for a given device. This requires incorporating data from psychophysical functions into the standardization procedure.

We emphasize that the objective of standardization <u>is</u> <u>not</u> to optimize image perception, but rather to equate, or make standard, image perception as defined by certain psychophysical tasks. The selection and measurement of the defining psychophysical tasks may vary depending on the type and format of images used, but the standardization methodology we outline in this paper readily accomodates these variations.

<u>Perceptual</u> <u>Linearization.</u> Based on considerations of the medical images with which we deal and what appear to be important psychophysical tasks in their utilization, we have developed a specific standardization protocol. We propose that equal changes between scale indicator intensities in the image should be displayed so as to appear equally discriminable. Thus, the function relating perceptual discriminability (for luminance) to scale indicator intensity would be described by a horizontal line. We refer to such a display as "perceptually linear" or "linearized".

To achieve this relationship between scale indicator intensities and perceived intensities, we must consider two functions: 1) for the display system, luminance or displayed intensity as a function of scale indicator intensity and 2) for the human observer, perceptual sensitivity to luminance differences as a function of luminance.

In this paper, we will describe the methods used for generating these two functions and their incorporation into a display system to make it linearized. The linearization process has been developed into a series of computer programs which can be readily adopted. Our psychophysical measurements have been incorporated into these programs so that implemention of the linearization processs only requires obtaining photometric measurements of the display system (i.e., measurements of displayed intensity vs scale indicator intensity). With slight modifications, the standardization technique described can accomodate other relationships (than perceptual linearization) between scale indicator intensity and perceived intensity, such as having equal relative changes in intensity appear equal.

CHARACTERIZATION OF DISPLAY MONITOR

The function relating digital driving level to photometric luminance of a monitor is called the greyscale. We measure the greyscale with the aid of an interactive computer program termed AUTOCAL. AUTOCAL sequentially displays images of uniform intensity, the order of (image intensity) presentations being specified by the user before initiating the program. Upon display of each intensity, the user makes a photometric measurement, enters this value into the program, and prompts for display of the next image. We set the program up to step through every driving level, and alternately run ascending and descending series, measuring luminance in footlamberts (fL) with a photometer. Measurements from each run are stored in separate files. The program AVG reads in a series of AUTOCAL files and outputs an AUTOCAL.AVG file which contains, for each driving level, the arithmetic average, standard deviation, and percent standard deviation of the intensity measurements made.

It is important that the greyscale lookup table addressed and used with AUTOCAL be clearly identified. The simplest strategy is to use the greyscale specified by the identity function. If a different greyscale is used, it is necessary to calculate and correct for differences between it and the identity function when constructing the linearized greyscale.

PSYCHOPHYSICAL MEASUREMENTS FOR LINEARIZATION

To produce a linearized greyscale as defined above, we must characterize the ability of the human observer to discriminate monochromatic luminances as being different intensities from each other. Numerous studies report that such discrimination thresholds are influenced by various stimulus parameters such as contrast (e.g., Legge and Kersten, 1983), stimulus size (Shapley, 1974), presence of sharp vs blurred edges (Campbell,Johnstone, and Ross, 1981), and presence of adjacent contours (van der Wildt and Waarts, 1983). Medical images are highly complex visual stimuli in that they contain multidimensional variations of these stimulus parameters. We would expect, therefore, that a fixed difference in luminance which is detectable when present in one region of a medical image may not be detectable when present in another region which contains, for example, nearby structures which may have an inhibitory effect on luminance discrimination. The local structure of images is an interesting issue with important implications for optimizing perceptiblity of images, and we plan to explore these issues in the future. Some preliminary studies we have made indicate that the effects of structure on linearization is a second order effect.

Our linearization transformation has been derived from luminance discrimination threshold measurements for very simple stimuli as described below. We measured sensitivity to luminance differences as a function of luminance by conducting psychophysical experiments for about 30 luminances spanning the range from black (0.002cd/m^2) to white (514 cd/m^2). Two procedures were employed, both based on signal detection methodology. One was a rating experiment and required subjects to rate their confidence that a signal was present on a trial. In the second, a two-alternative, forced-choice (2afc) procedure, each trial consisted of two temporally separated stimulus presentations. In one interval, but not the other, a signal was added to the stimulus. The subject's task was to choose whether the first or second interval contained the signal.

The stimuli were two squares, 1.5 cm on a side arranged one above the other, separated by 1 cm and located at the center of the video screen. The top square was maintained at a standard luminance (i.e., represents noise) for all trials of an experimental run. The luminance of the bottom, or comparison, square could be one of three possible levels. Relative to the top square, it could be equal in luminance, increased in luminance (i.e., contain a signal), but by slightly less than one jnd (as estimated from pilot data), or increased in luminance by slightly more than one jnd. To control for visual adaptation to the average display luminance level, the background was set so that the luminance from a field surrounding the test objects, corresponding to approximately 15 degrees of arc at the normal viewing distance of 62 cm, was 8.6 cd/m^2.

After a training period during which the subject was provided feedback with practice trials, an experimental session consisting of 120 trials without feedback, was run. For the rating experiment the observer was asked for each trial to estimate his confidence that the bottom square was brighter than the top square. For each increment in intensity (delta I) tested we obtained a true positive rate (i.e., the probability of correctly detecting the signal or luminance difference) and a false positive rate (the probability that the observer will detect a signal when none exists). For the two-alternative forced-choice paradigm, the observer's task was to choose the interval in which the comparison square was of increased intensity. We measured percent correct as a function of delta I.

Based on assumptions that the noise and signal-plus-noise distributions are represented by gaussian distributions with equal sigmas, we use a maximum-liklihood estimation program (Dorfman and Alf, 1969; modified by Metz and Kronman, 1980) to generate ROC curves for each delta I tested. Sensitivity is given by the value d', a parameter of the curve in ROC space. For any luminance, we define the luminance difference threshold (or jnd) as the value which has a d' of 1.645. For the rating experiments this corresponds to the delta I which would yield 50% true positive and 5% false positive; for the 2afc experiments it is the value associated with a 87.8% correct level of performance (Green and Swets, 1974).

Most of the data were obtained with the 2afc paradigm since, for the observer, it is easier than is the rating paradigm. The results, however, were independent of the method used and have therefore been pooled. Figure 2 summarizes these results. Jnd values, measured in cd/m^2, are the mean values for six observers. Not all observers participated in all conditions;

---

insert figure 2 about here

---

for each condition in which an observer ran, a single jnd has been calculated. When an observer ran more than one experimental session with a standard stimulus, the data were pooled to calculate a single jnd. The error bars shown on the plot represent the standard deviation of the mean for all measurements made at each point.

CONSTRUCTING A LINEARIZED GREYSCALE

Two computer programs are used to generate a greyscale. The first program, PPLIN, reads in the empirically acquired values which characterize the monitor and those which characterize the observer's visual sensitivity. The second program, COMTAB, uses the output of PPLIN to generate a video lookup table which, when installed in the display device, will produce a linearized greyscale. A slight variation, allowing users to enter psychophysical data other than ours, entails substituting a similar program JNDINTERP for the program PPLIN.

PPLIN

The psychophysical data, expressed photometrically, can be translated, for any display system, into digital driving levels (dl) to describe sensitivity for images displayed on that system. The program which performs this transformation is PPLIN. PPLIN has two inputs: 1) the AUTOCAL.AVG file for a specific display which gives the average photometric reading for each driving level, and 2) a table of the jnd data. For intensities subsumed within the photometric intensity range of the display, PPLIN translates the jnd data from photometric values into digital driving level values.

---

insert fig 3 about here

Figure 3 is a plot of the output from PPLIN and shows jnds as a function of stimulus intensity in units of driving levels (dl) for one particular monitor when its greyscale is defined by the identity function. These data are calculated using a piecewise multiple linear interpolation so that a jnd is calculated for every digital driving level of the display device (Pizer, 1981):

$$\mathrm{jnd}(y) = \mathrm{jnd}(i-1) + x(\mathrm{jnd}(i) - \mathrm{jnd}(i-1)) \quad \text{if } y \quad [i-1,i] \tag{1}$$

The resulting function, which we term the jnd curve, is shown in Fig. 3. This shows for every driving level of a display device that amount of change, measured in driving levels, which produces a just noticeable difference in intensity for our stimulus configuration. The values are stored in a table.

COMTAB

The program COMTAB uses the jnd curve for a particular monitor to produce, for that monitor, a LUT. This LUT generates a greyscale for which the jnd, measured for each driving level of the display, is represented by a constant number of driving levels.

The first step of COMTAB entails using the jnd curve to generate a curve summarizing the perceptual effect of display and observr. This curve is obtained by integrating infinitesimal increments of equal perceptibility up to each driving level. This gives what we term a perceptibility rank (P) for each dl, relative to 0 dl. Driving level 0 is given a perceptibility rank of 0. The rank for any other dl is then essentially determined by counting the discrimination steps (i.e., jnds) going from 0 to that dl. For example, the perceptibility rank of 70 can be approximated as 18 from the table by counting that there are 18 steps of one jnd each from 0 to 70, that is: one jnd from 0 to 7, from 7 to 10, 10 to 15, 15 to 27, 27 to 28, 28 to 29... 63 to 70. Using a limiting process that accumulates infinitesimal jnd fractions (Pizer, 1981), we obtain the equation below where P is the perceptibility rank and i is intensity in driving levels

$$P(i) = \int_{i_{min}}^{i} (jnd'(y) \;/\; (jnd(y)\ \ln(1+jnd'(y)))\ )\ dy \qquad (2)$$

Note that if the jnd curve is constant, the function P(i) is linear.

Figure 4 shows the relationship between the driving level scale and perceptibility rank for a specific display device. The slope of this function varies, providing a graphic depiction of nonuniform visual sensitivity to luminance changes generated by equal increments of the identity function greyscale; the steeper the slope, the lower the sensitivity.

---

insert fig 4 about here

The perceptibility rank for i max (the driving level of maximum intensity) is of special importance and defines the value of the perceptual dynamic range (PDR) of a display. The PDR is the number of jnds measured contiguously across a display greyscale.

The effect of COMTAB is to linearly rescale the jnds axis (from 0 to PDR) to dls (from 0 to dlmax) and then to invert the resulting function. Applying this inverse function as a LUT in the display device counteracts the effect of the display and observer so as to obtain an overall effect that is linear.

The PDR value, along with the value max dl, is used in the next step of COMTAB to rescale the P function ordinate into driving levels (scale indicator intensities) in such a manner that the dls are evenly distributed across jnds. This rescaling is effected by multiplying the P value for each dl of the identity function greyscale by the ratio (max dl - 1)/PDR.

---

insert fig 5 about here

In Figure 5, the function has been inverted and units along the abscissa are evenly spaced in the rescaled units of dls. This is the function which, when implemented in the display device by installation of a lookup table, will produce a linearized greyscale. All integer values along the abscissa are used as LUT addresses; the LUT values for these addresses are obtained from the ordinate of the function using a piecewise linear interpolation, followed by rounding to integers.

RESULTS

The linearization function for several devices has been calculated and applied to a number of images. Figure 6 is a photograph of a chest CT scan that has been contrast enhanced using a contrast limited adaptive histogram equalization program (Pizer, Austin, Perry, Safrit, and Zimmerman, 1986). The image is displayed on both an unlinearized and a linearized monitor. It is evident that the change made by linearization is quite substantial.

---

insert fig 6 about here

A linearized mapping, when compared to an unlinearized one (that is, one which uses an identity mapping from scale indicator intensity to device driving intensity), generally produces better separation in the lower intensity ranges. It should be emphasized, however, that the goal of standardization is not to obtain image improvement but to provide a standard baseline for the operating characteristics of the monitor from which the affects of the image processing can be correctly assessed and legitimate comparisons can be made between monitors. We propose that linearization is a means for the necessary preliminary step to sensible comparison of imaging devices and to consistent design of intensity mappings suited for a particular image and viewing context.

DISCUSSION

This paper presents a method for perceptual standardization of display devices, and a practical implementation is outlined. We have presented our method for measuring just noticeable differences (jnd) as a function of luminance. Our observer data were obtained specifically to address the presentation of medical images, and it may be that our particular experimental protocol is not the choice for other applications. However, the method of standardization can easily incorporate any observer data set that the user would wish to use. The process would involve designing the appropriate observer experiment, obtaining just noticeable differences in photometric units (fL), measuring the luminance output of the video screen as a function of digital driving levels, using the default lookup table, and then to use the JNDINTERP program. The JNDINTERP program allows input of psychophysical data different from ours.

A number of technical details merit comment. For example, when making photometric measurements of the video screen over the range of luminance levels, we find it useful to make maesurements at every driving level, i.e., 0 to 255 for our system. We make several sets of measurements and calculate the average luminance at each driving level. A series of such measurements provides us with a record of reproducibility of the measurements and stability of the monitor. We have also found that once we select the contrast and brightness settings for a monitor, measurement of the maximum luminance and minor adjustment of the contrast control is adequate to maintain calibration of the monitor.

Some spatial and temporal variations in luminance of video monitors that one must consider when standardizing them is that luminance of the phosphor changes with warm-up time. We have found that at least 30 minutes is necessary for stabilization. The uniformity of screen luminance is ±10% or worse. We have elected to conduct all our photometric measurement at the center of the screen. Of course a better method, which we are currently implementing, combines one image plane (1024 x 1024 x8) which has the correction factors for the screen nonuniformity with the image plane that has the display.

A method that accomplishes standardization depends on a number of parameters, e.g. image structure, variation of contrast within the image, influence fo ambient light, and the differences between observers. Some of these we have investigated [Rogers, Johnston, and Pizer, 1987; Johnston, Zimmerman, and Pizer, 1985]. While the application of a single global standardization for the whole image is clearly useful, global linearization may not be the standardization method of choice for all display situations. We are continuing research in this area to investigate how jnds are affected by target surround contrast, spatial structure of the target and spatial structure and luminance of the surround. Matching perceptual linearization to image structure could lead to distinct improvements in diagnostic performance in medical images.

Finally, we state again that the goal of standardization is not specifically to obtain image improvement, but rather to provide a standard baseline for the operating characteristics of the monitor from which the effects of image processing can be properly assessed and legitimate comparisons can be made between monitors. Our method of linearization provides a means for the necessary step to sensible comparison of imaging devices and for consistent design of intensity mappings suited for a particular image and viewing context.

ACKNOWLEDGEMEMTS

This research has been supported in part by NIH Grant 1Ro1-CA39059 and a NATO Collaborative Research Grant No RG 85/0121.

REFERENCES

Campbell, F.W., Johnstone, J.R., and Ross, J. An Explanation for the Visibility of Low-frequency gratings. *Vision Research*,*21*, 1981, 723-730.

Dorfman, D.D. and Alf Jr.,E. Maximum Likelihood Estimation of Parameters of Signal Detection Theory and Detection of Confidence Intervals-Ratings-Method. *Journal Mathematical Psychology*, 1969, *G*, 487-496.

Green, D.M. and Swets, J.A. *Signal Detection Theory and Psychophysics.* New York: Robert E. Krieger Publishing Co., 1974.

Johnston, R.E., Zimmerman, J.B., Rogers, D.C., and Pizer, S.M. Perceptual Standardization.*Picture Archiving and Communications System (PACS IV) for Medical Applications*. 1985, SPIE *536*, 44-49.

Legge, G.E. and Kersten, D. Light and Dark Bars: Contrast Discrimination.*Vision Research*,*5*, 1983, 473-483.

Metz, C.E. and Kronman, H.B. Statistical Significance Tests for Binormal ROC Curves. *J of Mathmatical Psychology*, 1980, *22*, *3*, 218-243.

Pizer, S.M. Intensity Mappings to Linearize Display Devices. *Computer Graphics and Image Processing*, 1981,*17*, 262-268.

Pizer,S.M., Austin,J.D., Perry, J.R., Safrit, H.D., and Zimmerman, J.B. Adaptive Histogram Equalization for Automatic Contrast Enhancement of Medical Images.Application of Optical Instrumentation in Medicine XIV, 1986,SPIE 626, 242-250.

Rogers, D.C., Johnston, R.E., and Pizer, S.M. Effect of Ambient Light on Electronically Displayed Medical Images As Measured by Luminance Discrimination Thresholds. J. Op. Soc. Am 4, 5, May 1987.

Shapley, R. Gaussian Bars and Rectangular Bars: The Influence of Width and Gradient on Visibility. Vision Research,14, 1974, 1457-1461.

Van der Wildt, G.J., and Waarts, R.G. Contrast Detection and its Dependence on The Presence of Edges and Lines in The Stimulus Field. Vision Reserach, 23, 1983, 821-830.

LIST OF FIGURES

U

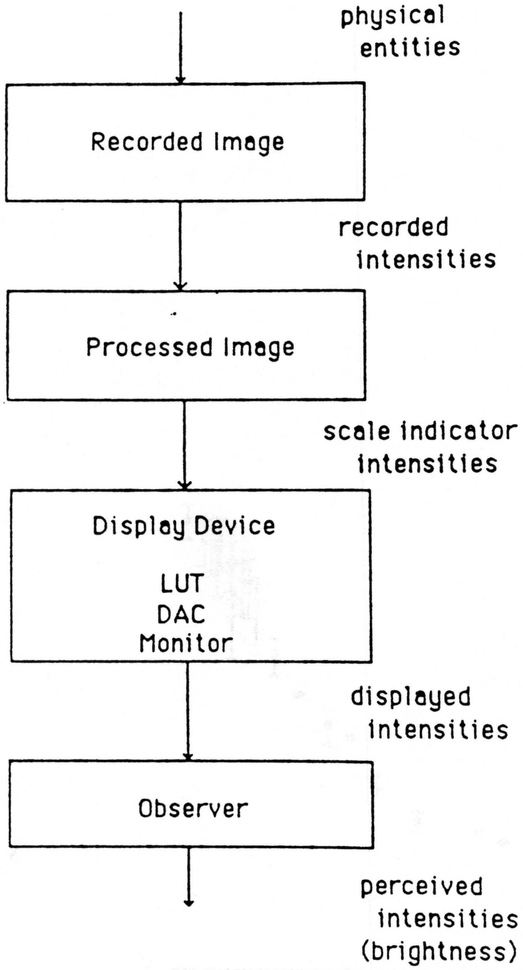

physical entities
Recorded Image
recorded intensities
Processed Image
scale indicator intensities
Display Device
LUT
DAC
Monitor
displayed intensities
Observer
perceived intensities (brightness)

Fig 1

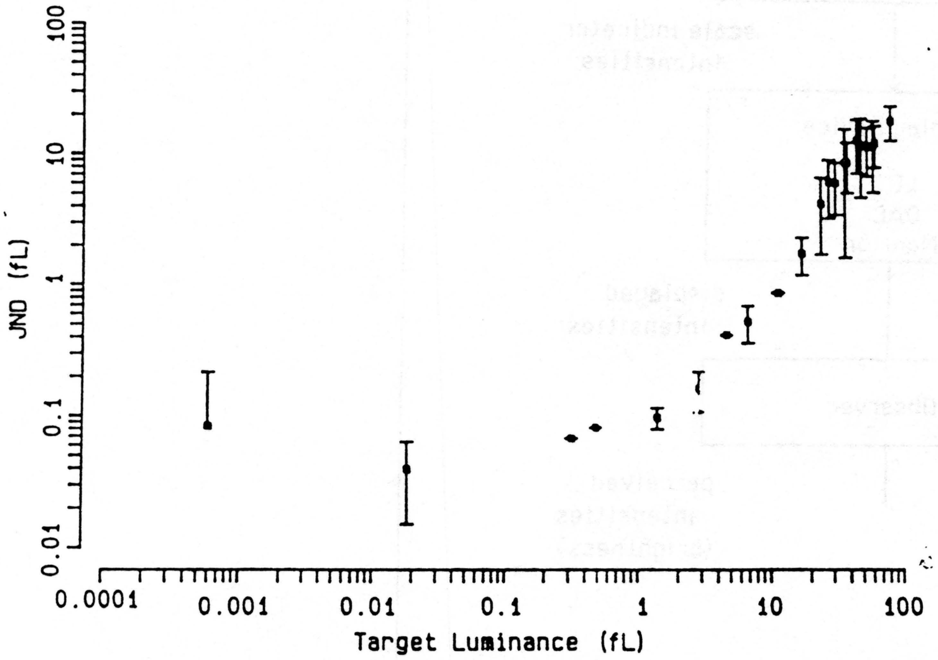

JND (fL)
100
10
1
0.1
0.01
0.0001
0.001
0.01
0.1
1
10
100
Target Luminance (fL)

Fig 2

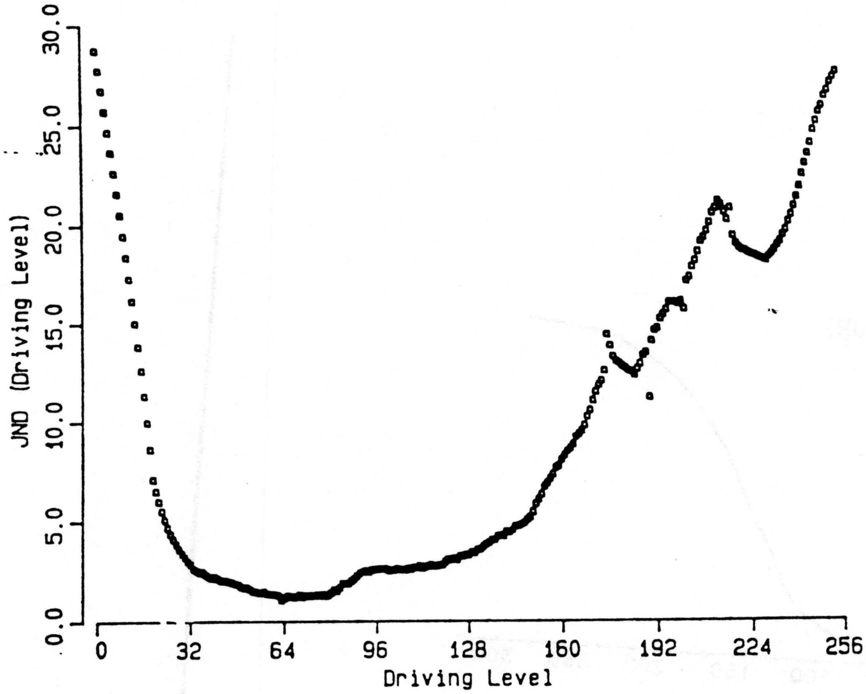

JND (Driving Level)
30.0
25.0
20.0
15.0
10.0
5.0
0.0
0
32
64
96
128
160
192
224
256
Driving Level

Fig 3

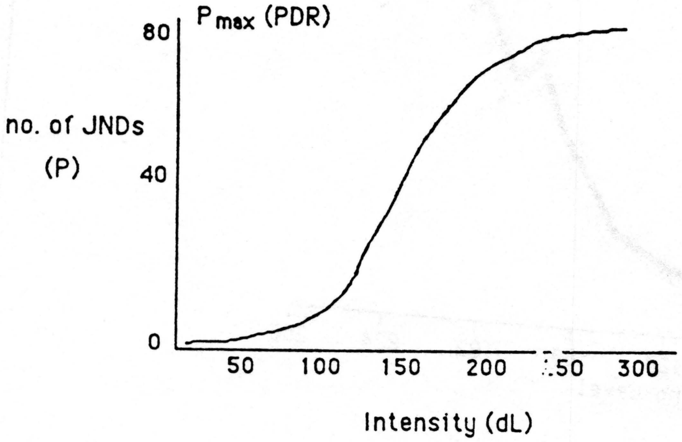


Fig 4

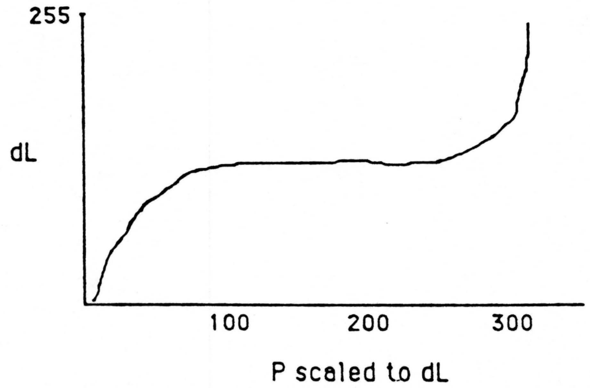


Fig 5

Top ↑

Fig 6