LUMINANCE DESIGN : A SIMULATION USING COLOUR TELEVISION

by

D. J. Gilderdale B.Sc.

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy of the Council for National Academic Awards.

This work was carried out in the School of Electrical and Electronic Engineering, Plymouth Polytechnic.

February 1982

ABSTRACT

Luminance design : a simulation using colour television.

by D. J. Gilderdale

A successful lighting design usually results from the skill of the designer in applying professional experience. However, successful designs have been achieved using numerical prediction. It is probable that a blend of both these elements will give the optimum result. Whatever the design approach, the end product will be judged, at least in part, on its aesthetic merits. The first chapter of this thesis introduces the possibility of using a digital computer in conjunction with a colour television monitor to calculate and display the luminance distribution in a lighted room; a system which may offer advantages both for the experienced designer and the student of lighting design. The display system is described briefly, along with some possible shortcomings. An account is given of the methods used for inter-reflection calculation. These inter-reflection calculations are then developed to include colour and techniques of photometric and colorimetric measurement with reference to the television display. A complete description of the display system hardware is also given. This display system as initially designed uses chromaticity as the criterion for colour reproduction. The shortcomings of this approach are discussed. Techniques for perceived colour measurement are described and the results presented for the colour perceived from some simple display images. The possibility of perceived colour prediction is examined and measured colours are compared with those predicted by a non-linear model. Finally, the applications of the display are discussed, both in an educational and design context. Some possible developments and improvements are also outlined.

Chapter 1

INTRODUCTION

During the past twenty years much research activity has been devoted to the problem of producing optimum light distributions in buildings, particularly those providing a work environment. If the ceiling is too bright it may be distracting; if too dark it may be oppressive. Ideally, the walls should be slightly darker than the visual task. The distribution of illuminance depends, in a complex but predictable manner, on the position and intensity distribution of the lighting equipment and on the shape and reflectances of the interior surfaces.

A design process based on the prediction of illuminance, although mathematically straightforward, involves rather lengthy and tedious calculations which are best entrusted to an electronic computer. This arithmetic load is perhaps the main reason why luminance design has been more often preached than practised.

In Figure 1.1 is shown a system for simplifying the design procedure whilst at the same time producing a simulated display of the luminance distribution using television. A designer is presented with the relative luminances produced by a chosen lighting system and surface reflectance distribution. The system employs colour television, enabling the designer to try out different coloured surfaces at the same time as he examines different lighting arrangements, providing both quantitative and qualitative information.

The speed of calculation and immediacy of the output display allows a designer to make rapid changes so that his design may be optimised. As indicated in Figure 1.1, the system may be divided into three sub-sections, the software lighting model, the image generation hardware and the display viewer. These will be introduced separately.

1.1 Calculation of the light distribution.

Calculation of the lighting distribution is carried out by an electronic computer which may be of analogue or digital type. The tedious arithmetic operations are therefore performed very quickly, leaving the operator free to concentrate on the detail of the lighting scheme.

In order to provide an element of computer-aided design, the present system uses a digital computer which, used with a high level language, allows for a convenient dialogue between machine and operator. This interaction is via a conventional teletypewriter or V.D.U. The system is not restricted to one type of digital computer and may be interfaced with microprocessors, mini or even mainframe computers.

Data input via the keyboard is controlled by software which guides the operator through a series of questions regarding the specification of the lighting scheme before going on to process the final luminance distribution. Having calculated and displayed this lighting scheme, the software prints out various options which the operator may choose to follow; he may wish to change the colour and/or reflectance of one or more of the surfaces or, alternatively, he may wish to change the

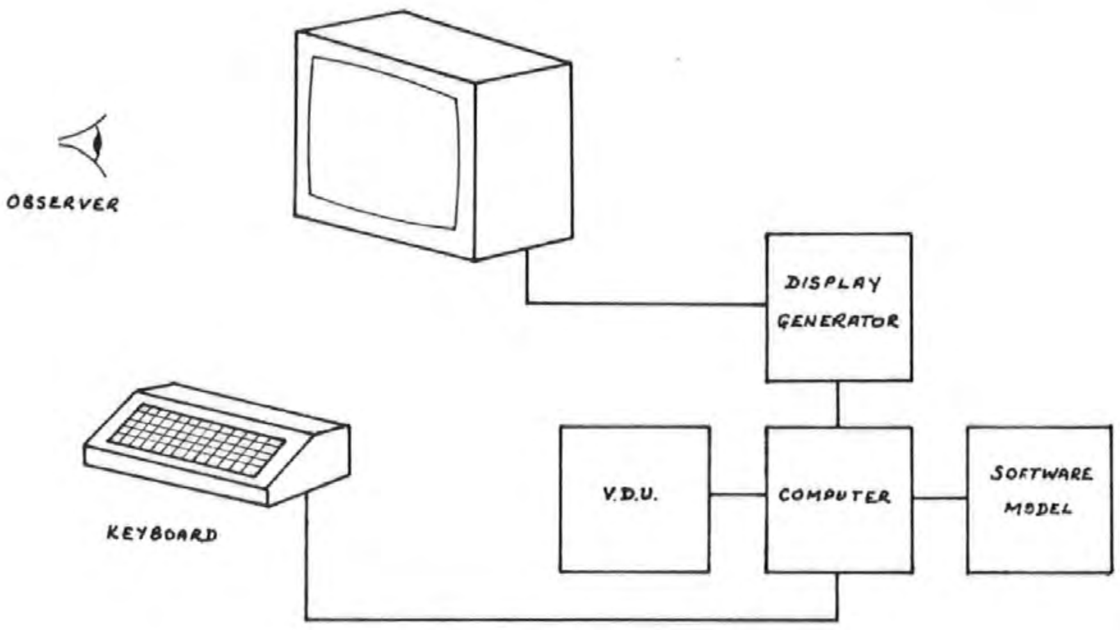


Fig. 1.1 The display system.

nature of the lighting equipment. A further option would be that the operator has found an optimum design, in which case the programme terminates with a printout of the complete lighting/colour specification.

The aim is that with this system, it should be possible for a person with little knowledge of lighting engineering to learn about lighting design by means of the numerical and visual feedback provided.

A further feature, handled by the software, involves the lighting design limits for work environments as specified in the Illuminating Engineering Society 'Code for Interior Lighting' 1977. Here are quoted recommended ranges of reflectance and relative illuminance for the surfaces of work spaces. If required, the software can automatically check at the end of each design run that the illuminances fall within the specified limits and advise the operator appropriately.

1.2 The image generation system.

The impression of the perspective view of a room interior is produced electronically by a digital system which is synchronised with the raster scan of the television display. As with the computation, the same function could have been performed by analogue circuitry. However, the advantages of digital circuitry in terms of stability, reliability and cost, were overwhelming in the choice of circuit design.

As with all short-persistence raster scan displays, it is necessary to refresh the image at a repetition frequency sufficiently high to avoid subjective flicker. Thus the image information for one complete frame is stored electronically by the display generator. Data determining the image vanishing point and the colour appearance of each of the five displayed room surfaces is output from the computer via a serial data link, being transferred sequentially into memories. The display and computer run independently and asynchronously.

Perspective data is used in digital form whilst the colour information is converted into the analogue form required by the television monitor before being switched by synchronous logic circuits to produce a video signal representing the room interior.

1.3 Image perception and the display.

The relationship between the perceived quantity, brightness, and the luminance of a stimulus is known for foveal stimuli with a dark surround. In complex images this relationship is less predictable, as is that between chromaticity and perceived colour. For this reason luminance level has been brought into question as an appropriate design criterion. A television display simulation may offer the possibility of recreating the same perception of brightness and colour as would be produced by the original scene and hence minimise one of the shortcomings of luminance design.

The viewer forms the final and least predictable link in a television display system. The viewer's response is a function of his viewing position and ambient lighting conditions, as well as the display stimuli. The later sections of this thesis are devoted to a preliminary study of colour perception from a display screen. If perceived colour can be predicted, it should be possible to design a display system using colour sensation rather than chromaticity as an optimum display criterion.

Chapter 2

THE PREDICTION OF ILLUMINANCE

This chapter gives a brief account of the development of luminance design. The derivation and solution of the equations for particular models are described with reference to the display system discussed in chapter 3.

2.1 Lighting prediction in buildings - a brief history.

Early lighting designers were preoccupied with the efficient use of what little light was available. In 1920, the Americans Harrison and Anderson[1] published a design technique based on the concept of flux transfer, which predicted the proportion of the luminaire flux output which intercepted a work plane nominally 0.85 m above the floor. This fraction of useful flux was called the utilance. The Harrison-Anderson method adopted an empirical approach to the problem of inter-reflection. The utilances for five luminances in a range of square rooms of varying proportions and surface reflectances were catalogued, allowing the designer to approximate for particular design conditions. Rectangular rooms were described in terms of two square rooms having the length and width as their respective dimensions. This design technique, known as the 'Lumen' method, gained widespread popularity.

As new luminaires became available and power more plentiful, the shortcomings of the pragmatic Harrison-Anderson method became apparent.

A more specific approach to inter-reflection was needed, giving information about illumination not confined to the work plane.

In 1862, Stokes[2] had developed an exact solution to the problem of inter-reflection between two parallel plates, analogous to a room of infinite length and width, by adding successive reflections. Using a similar iterative approach to inter-reflection Sumpner (1892)[3] was able to predict illuminance for the more practical case of a non-infinite interior. In 1920, Ulbricht[4] solved the same problem using a more elegant technique based on the concept of energy conservation. The steady state model treated the interior as equivalent to an integrating sphere as used in photometry, both surface reflectance and illuminance being assigned average values. Attempts were made by Phillips (1956)[5] and Cuttle (1973)[6] to apply this method to more practical lighting situations, not, however, without approximation. The direct component of illuminance may be accurately specified but that due to inter-reflection is assumed to be the same average value for each surface. In conditions dominated by direct lighting this expedient can produce useful results. The method proposed by Cuttle is discussed further in a later section of this chapter.

An exact solution of the inter-reflection problem requires an accurate specification of both direct and indirect components of illuminance. The steady state problem is described by an integral equation, for which an analytic solution is possible for a simple geometry such as a sphere. Moon and Spencer (1946)[7] showed that a solution of the equation was possible for geometries of practical significance, such as a rectangular room, but only by introducing approximations.

An alternative solution to the integral equation was first suggested by Yamauti[8] in 1926 who used a finite difference approximation which is solved numerically. This technique was later revived by Dourgnon in 1953.* In the following year, Waldram[9] proposed an alternative design procedure which was to stimulate considerable interest. This designed appearance method involved the calculation of the distribution of direct illuminance required to produce a pre-defined luminance distribution. Dourgnon[10] later showed the Waldram method to be an inverse statement of the finite difference equations used to calculate total luminance from direct illuminance. In 1959, O'Brien and Howard[11] demonstrated that the finite difference approach was more accurate than the Moon and Spencer approximate integral equation, even when relatively few elements were used. O'Brien (1959)[12] rearranged the finite difference equations into a form suitable for solution by analogue computer to produce, in principle, instantaneous results. The steady state approach was also developed by Lynes (1959)[13] who showed that the effects of specular reflection, variable wall luminance, interior obstructions and colour could also be included. The inclusion of colour is further discussed in chapter 3 of this thesis.

An unsolved problem concerned the relationship between the direct illuminance distribution and a practical lighting installation. In 1961, the Illuminating Engineering Society published a report describing a routine method for predicting the pattern of illuminance, known as the British Zonal (BZ) method.

The BZ system includes a simple method for classifying luminaires from which the direct illuminance distribution may be calculated

* He had developed this technique independently in 1928.

for a given interior, This system, described in I.E.S Technical Report No. 2[14], is used in the design method described later in this chapter.

2.2 Derivation of the equations.

More than thirty years ago, Yamauti (1926)[8] showed that the luminance L_i of a surface i could be expressed by the equation for conservation of flux as

$$L_i = L_{oi} + \rho_i E_i \qquad 2.1$$

where L_{oi} is the surface luminance due to direct illumination

ρ_i is the surface reflectance

E_i is the illuminance due to inter-reflection

The problem now concerns the inter-reflected component E_i and its relationship with the geometry of the room and its lighting system.

Consider two infinitesimal planar surface elements δA_i and δA_n as shown below which may have arbitrary position and orientation.

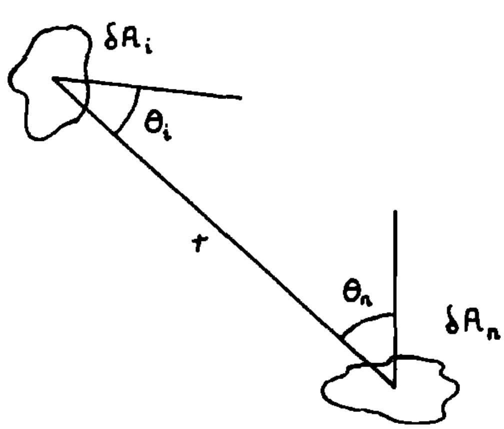


Figure 2.01

The illuminance of surface i from any other surface n may be expressed as

$$E_i = \frac{1}{\pi A_i} \int_{A_i} \int_{A_n} \frac{L_n \cos\theta_i \cos\theta_n}{r_{i-n}^2} \, dA_n \, dA_i \qquad 2.2$$

(for development of the derivation of this equation see Appendix 2A)

where: A_i and A_n are the respective surface areas

r_{i-n} is the distance between the two elements

L_n is the luminance of surface n

θ_i and θ_n are the angles formed between the normals to dA_i and dA_n and the line r_{i-n} connecting these elemental areas.

Thus, using the above two equations it should in principle be possible to calculate the total luminance of a surface resulting from direct and inter-reflected illumination. These equations may be solved directly for a very simple geometry, e.g. a sphere (see Moon 1936)[15]. Unfortunately, more practical geometries (at least from the lighting viewpoint) such as a rectangular enclosure give rise to integrals which do not give an analytic solution.

2.3 An approximation using the exponential 'kernel'.

One further step from the sphere in geometrical complexity, e.g. the case of a cylinder, produces an integral equation which is not directly soluble. In this case the 'kernel' of the integral in equation 2.2 has a form which prevents direct integration.

One approach to this problem for the case of a cylinder was suggested by Buckley (1927)[16] and further developed by Moon and Spencer (1946)[7]. This consists of approximating the kernel by an exponential function which is related to the relative geometry of the surface elements. Having made this approximation the equation may be solved analytically. Moon and Spencer further develop this exponential approach for the case of a rectangular enclosure, this having more practical relevance to lighting calculations. These methods are discussed more fully in Appendix 2B.

On the assumption that any solution is better than no solution this approximate method has proved useful; however, the accuracy is limited and is a function of the relative dimensions of the rectangular enclosure.

2.4 A finite difference approach to the inter-reflection equation.

An approach often used to solve integral equations not having a direct analytic solution is to apply a 'finite difference' approximation. This method was first suggested by Yamauti in 1926[8] but its application to the numerical solution of the above integral has not been developed until comparatively recently. Work by O'Brien and Howard (1959)[11] has shown that this method can give a significant improvement in accuracy over the 'exponential kernel' approximation method.

With the finite difference approach, the system is divided into small elements which are assumed to have uniform characteristics. The accuracy of the method improves as the number of subdivisions or 'finite elements' is increased.

In the lighting application, the variables of equation 2.1 are postulated to exhibit constant or uniform reflectance ρ and luminance L over some finite region. Using this approximation the functions ρ and L may be separated from the integral leaving only a geometrical function to be integrated over the space co-ordinates specified. For most common geometries, this function has been integrated and is available as a standard form (Hamilton and Morgan, 1952)[17]. Lynes (1959)[13] uses charts to derive these functions as part of a design procedure. With the luxury of a digital computer, however, these functions may be calculated directly.

The functions which define the geometrical relationship between any two surface elements, are called 'form factors'. For elements one and two, for example, the form factor f_{12} is equivalent to the fraction of the total flux emanating from element 1 which is intercepted by element 2. For the integral equation (2.2) above, the equivalent finite difference equations, which describe regions A_1, A_2 ....A_n each having constant ρ and L values, may be written:-

$$L_1 = L_{01} + \rho_1 \left[f_{12}L_2 + f_{13}L_3 + \ldots . f_{1n}L_n \right]$$

$$L_2 = L_{02} + \rho_2 \left[f_{21}L_1 + f_{23}L_3 + \ldots . f_{2n}L_n \right]$$

$$\vdots \qquad\qquad \vdots$$

$$L_n = L_{0n} + \rho_n \left[\qquad + \qquad + \ldots . f_{nn-1}L_n \right] \qquad 2.3$$

The form factors f_{nn} do not appear here because the surface elements are assumed planar and therefore a surface cannot illuminate itself. Hence $f_{11} \; . \; . \; . \; . \; f_{nn} = 0$.

Thus, if the direct components of luminance and the reflection coefficients are known, the system of linear first order equations can, in principle, be solved to determine all the surface luminances.

The derivation of the form factors is described more fully in Appendix 2C.

2.5 Comparative accuracy of the approximation methods.

Sections 2.2, 2.3 and 2.4 show the equation of flux transfer to be fundamental to the lighting system but only directly soluble for certain simple shapes such as a sphere. For more practical cases, some form of approximation must be introduced. Here two possibilities remain:

(a) To evaluate the integral 'analytically' but make some approximation for the geometrical function, e.g. by an exponential function.

(b) To use a finite difference approximation, but then to use exact values for the resulting geometrical functions or form factors.

The work by O'Brien et al (1959)[11] has shown that when the same approximate values for the shape moduli or form factors are used, the two approaches give virtually identical results. When, however, exact values are substituted in the finite difference method the results from the two approaches differ noticeably. The same work also shows that the finite difference method with exact form factors is more accurate even when only a few finite elements are used and thus provides a powerful tool for inter-reflection analysis.

An interesting comparison for the exact and finite difference methods is given by O'Brien and Howard (1959)[11]. For the case of an infinitely long hallway Spencer (1958)[18] has derived an exact analytic solution to the Fredholm integral. O'Brien et al applied the finite difference approach to this case with the walls represented by one, two and four areas or equal sections between the floor and ceiling. Truncation errors for the case of four wall sections are sufficiently small to make further subdivisions unnecessary. The greatest error occurs for the wall luminance adjacent to the ceiling since it is in the region that the maximum gradient of luminance occurs.

2.6 Solution of the equations

The finite difference equations as described in section 2.4 describe an inter-reflecting system of n elements. The simplest representation of a rectangular room requires the solution of six linear first order equations in six unknowns. Various numerical methods (e.g. Gaussian elimination) exist for the solution of simultaneous equations by hand calculation. For six equations, the large amount of computation is best handled by an electronic computer. The performances of some typical numerical methods are compared in section 2.62.

2.61 An analogue model.

O'Brien (1960)[12] has suggested a technique allowing the finite difference form of the flux transfer equation to be solved using an electronic analogue computer. The equation for luminous flux transfer may be rewritten in matrix form:

$$\begin{bmatrix} 1/\rho_1 & -f_{12} & -f_{13} & -f_{14} & -f_{15} & -f_{16} \\ -f_{21} & 1/\rho_2 & -f_{23} & -f_{24} & -f_{25} & -f_{26} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ -f_{61} & -f_{62} & -f_{63} & -f_{64} & -f_{65} & 1/\rho_6 \end{bmatrix} \begin{bmatrix} L_1 \\ L_2 \\ \cdot \\ \cdot \\ \cdot \\ L_6 \end{bmatrix} = \begin{bmatrix} L_{01}/\rho_1 \\ L_{02}/\rho_2 \\ \cdot \\ \cdot \\ \cdot \\ L_{06}/\rho_6 \end{bmatrix} \qquad 2.4$$

By algebraic manipulation it is possible to rearrange this set of equations into the form of Kirchhoff's node equation:

$$\begin{aligned} &\frac{L_1 - L_2}{1/A_1 \cdot f_{12}} + \frac{L_1 - L_3}{1/A_1 \cdot f_{13}} + \cdots \frac{L_1 - L_6}{1/A_1 \cdot f_{16}} = \frac{(L_{01}/(1-\rho_1)) - L_1}{\rho_1/(1-\rho_1).A_1} \\ &\frac{L_1 - L_2}{1/A_1 \cdot f_{12}} + \frac{L_2 - L_3}{1/A_2 \cdot f_{23}} + \cdots \frac{L_2 - L_6}{1/A_2 \cdot f_{26}} = \frac{(L_{02}/(1-\rho_1)) - L_2}{\rho_2/(1-\rho_2).A_2} \\ &\cdot \\ &\cdot \\ &\cdot \\ &\frac{L_6 - L_1}{1/A_1 \cdot f_{16}} + \cdots\cdots \frac{L_6 - L_5}{1/A_6 \cdot f_{65}} = \frac{(L_{06}/(1-\rho_6)) - L_6}{\rho_6/(1-\rho_6).A_6} \end{aligned} \qquad 2.5$$

This equation is recognised as Kirchhoff's node equation if the luminances in the numerators of each fraction are regarded as potentials and the denominators, composed of geometry functions, are regarded as resistances or energy dissipators. Thus, it should be possible to construct an electrical resistance network whose nodes have potentials L_1, L_2 . . . . L_6 which are exact analogues of the surface luminances in the luminous flux transfer problem. The terms on the left hand side of the equation show that each node is connected to every other node by a path which contains a resistance $1/A_n f_{nm}$. The current through each branch of the network is analogous to the luminous flux flowing from one surface to another.

The terms on the right of equation 2.5 represent the sources of luminous flux which depend on the driving potential $L_{on}/(1 - \rho_n)$. Fig. 2.02 shows the resistive network required to represent an array of six equations.

In this resistive analogue, variable power supplies are used to set potentials proportional to $L_{on}/(1 - \rho_n)$ at each of the outer nodes as shown. For surfaces which are not directly illuminated, the corresponding input node is connected to ground potential. All electrical potentials are measured with respect to a common ground potential.

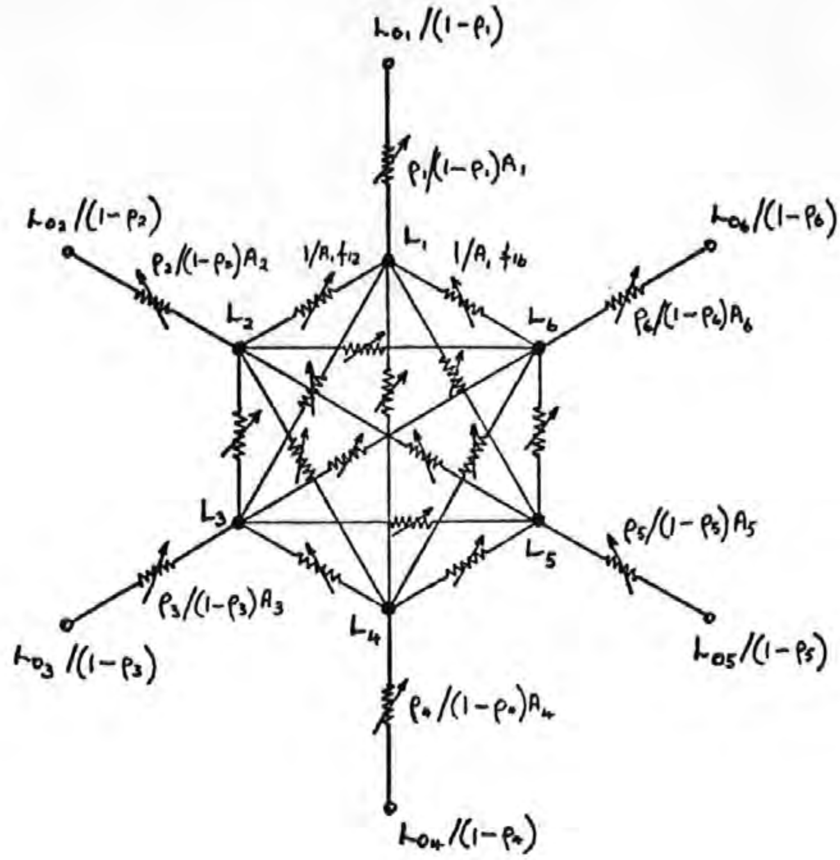


Figure 2.02

For this simple case of six surfaces it is evident that twenty-one precision variable resistances are required in addition to six power supplies which, depending on the particular problem conditions, may be required to supply or dissipate considerable amounts of power. In view of the special and rather expensive components needed for this circuit it was decided, at least in the first instance, to represent the equations on a conventional analogue computer using active components, since a suitable computer was readily available.

Fig 2.03 shows part of the circuit required to represent equations using an EAI TR48 analogue computer.

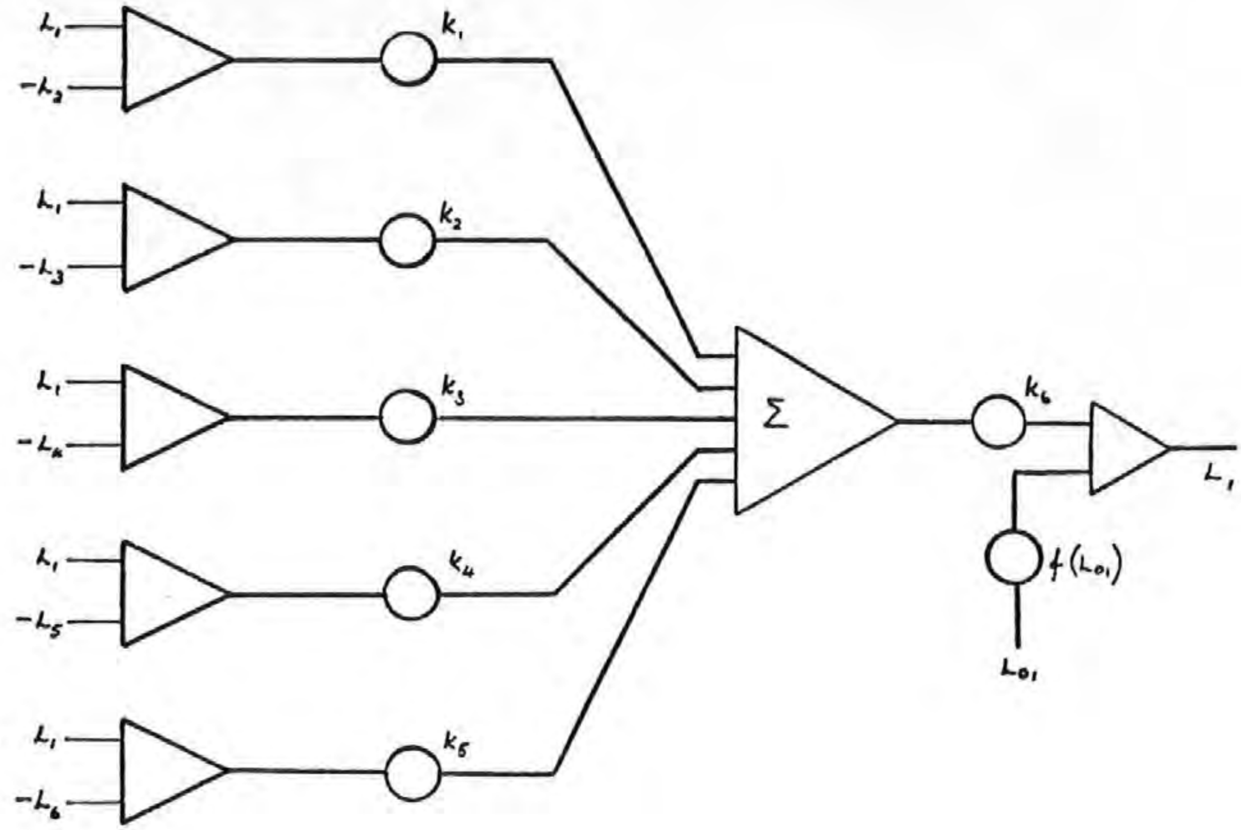


Figure 2.03

Each term on the left hand side of the Kirchhoff node equation is represented by a difference amplifier and a potentiometer. These terms, in addition to a driving term representing the direct illumination, are summed to yield the total luminance of one of the surfaces (in this case L_1).

For the rectangular room model six such circuits are interconnected to provide the solution $L_1 \rightarrow L_6$. Although the amount of wiring is large compared to the purely resistive model, this system has the advantage that standard analogue computer elements may be used and no circuit components are required to dissipate large amounts of power.

The advantage offered by both forms of analogue computer is that the input data having been set, the solution is produced almost instantaneously. Against this must be weighed the problem of wiring complexity and the formidable amount of potentiometer setting and calibration which must be carried out prior to obtaining a first solution. Subsequent solutions may require the adjustment of only one parameter, in which case results are obtained without delay.

2.62 Solution by digital computer.

As mentioned earlier, various methods have been developed for the solution of linear first order simultaneous equations by hand calculation. More recently many of these techniques have been adapted for use with a digital computer.

Used with a high level language, the digital computer may be conveniently programmed to provide a dialogue between machine and user. Thus it is possible to use a digital machine to control the input of data and thus guide a designer through the process of inputing a lighting system specification. In this respect the digital machine provides significant advantages compared to analogue computers.

Having input the relevant data, the problem of deriving a solution for the six simultaneous equations still remains. Four standard numerical methods have been tried, each capable of providing solutions of sufficient accuracy, but each having advantages for particular applications.

Table 1 lists the methods used and summarises their performances in solving equation 2.4.

Table 1

METHOD	Time in seconds to solve 6x6 matrix equation in 'FOCAL' PDP8/F	PDP11/10
Gauss Seidel iteration	24.5	21.5
Matrix inversion	15	11.5
Gaussian elimination	6	4.5
Crout's method	7	5.5

2.7 Representation of the lighting fittings - the BZ method.

The lighting model, as described in section 2.2 enables calculation of the surface luminances to be made, on the assumption that the direct component of illuminance for each surface is known. In a practical lighting system these direct components are related in some way to the distribution and intensity of luminous flux emitted by the light fittings.

In 1961, the Technical Committee of the Illuminating Engineering Society produced a report describing an approach to classification of lighting installations and calculating resultant luminances called the 'British Zonal' or BZ system. In the display system described in this report, all light fittings are assumed to be capable of classification using the BZ method.

2.71 Terminology.

Before attempting to describe the BZ classification it is useful to define some of the more important terms used.

Flux fraction.

The proportion of total flux emitted from a light fitting in the upper or lower hemisphere.

Flux fraction ratio (FFR)

The ratio of the upward flux to the downward flux from the fittings in an installation.

Direct ratio (DR)

The proportion of the total downward flux from a conventional installation of lighting fittings which is directly incident on the working plane. (This is a function of the relative geometry of the working plane and the light fitting).

Mounting height (Hm)

The height of the fittings above the working plane.

Suspension length (Hs)

The height of the ceiling above the fittings.

Spacing (S)

The distance between rows of fittings in square arrangement.

Room Index (k_r)

This is related to the shape of a rectangular interior according to the formula

$$k_r = \frac{LxW}{H_m(L+W)}$$

where L and W are respectively the length and width of the interior.

Ceiling cavity index (k_c)

An index relating to the shape of the cavity above the level of the lighting fittings in a rectangular interior according to the formula

$$k_c = \frac{LxW}{H_s(L+W)}$$

2.72 The BZ classification

The method of classification is based on the fact that for any given interior, the illumination reaching the work surface is related to the direct ratio (DR) of the lighting installation. The amount of downward flux which reaches the work surface directly can be found from knowledge of the direct ratio and the lighting input. The former is a function of room index and the layout of the lighting fittings.

As a convention, a regular square arrangement of fittings is assumed, at a uniform spacing/mounting height ratio (S/H_m) with one half of the normal spacing between the fittings and the walls. Such precise symmetry is unlikely to occur often in practice but small deviations from these conventional conditions may be compensated for by calculation.

In many applications, the spacing/mounting height ratio is determined by a need to meet a condition for uniformity of illumination. The 'Uniformity ratio' is defined as the ratio of the minimum to the maximum illuminance in the central area of the installation.

For most practical installations a uniformity ratio of 0.7 is recommended.

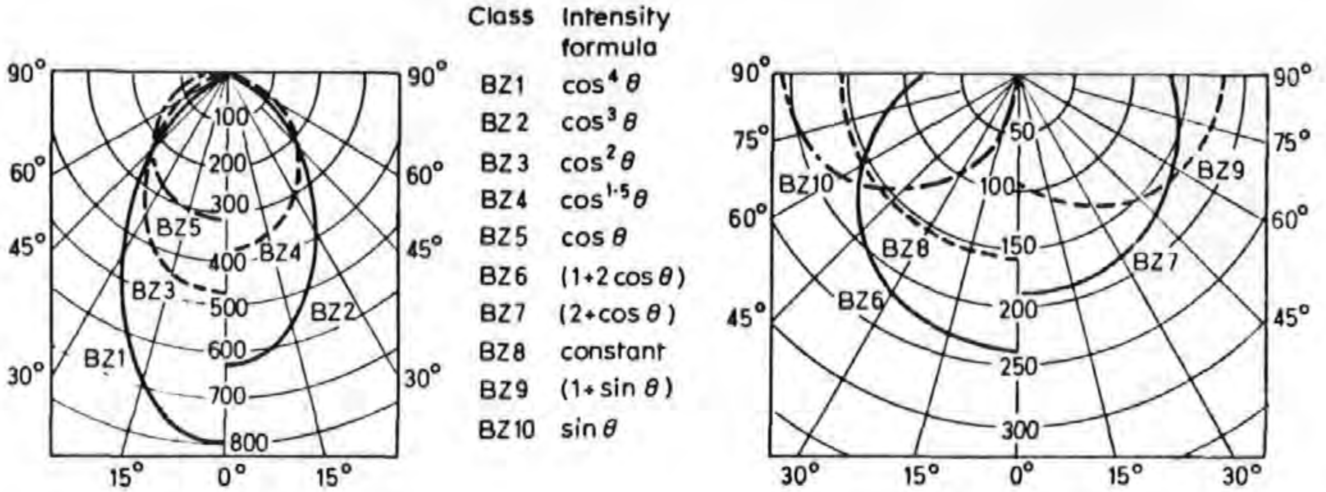


Fig. 2.04 Polar distributions for various BZ classifications.

The downward light distribution for an array of fittings may be described by plotting luminous intensity in polar form. Figure 2.04 shows ten theoretical polar distributions which cover a range of beam spreads from the narrowest to widest likely to be met in practice. These distributions are the basis for the BZ classification.

The designer selects a uniformity ratio from which may be obtained a value of spacing/height ratio for each polar distribution. From these S/H_m values it is possible to plot curves of direct ratio against room index, one for each polar distribution. These reference curves are shown in Figure 2.05 and enable the direct ratio to be obtained for any installation which may be classified using the BZ system.

For ceiling or near ceiling mounted fittings all the upward flux from the fittings will intercept the ceiling. This upward component is defined by the flux fraction ratio FFR. Thus if the values of direct ratio and flux fraction ratio are known it is possible to predict the distribution of direct light amongst the interior surfaces. Taking the simple case of an empty room (assuming the work plane to be the floor and the walls all have the same direct illuminance)

Direct floor illuminance $$\frac{\phi_L \cdot DR}{A_F\ (1+FFR)}$$

Direct ceiling illuminance $$\frac{\phi_L \cdot FFR}{A_F\ (1+FFR)}$$

Direct illuminance of any one wall $$\frac{\phi_L \cdot (1-DR)k_r}{2A_F\ (1+FFR)}$$

where: ϕ_L = Total flux output from lighting installation

DR = Direct ratio

FFR = Flux fraction ratio

A_F = Floor (ceiling) area

Thus, using the BZ system, it is possible, for a given lighting installation in a rectangular room of given dimensions, to calculate the direct components of luminance Lo_i as used in the finite difference model discussed in section 2.4. The BZ method provides an easy and rapid way of predicting direct lighting levels for a given room geometry and lighting installation.

To find direct ratio from the curves of Figure 2.05 it is necessary to decide on the room index value and the BZ classification.

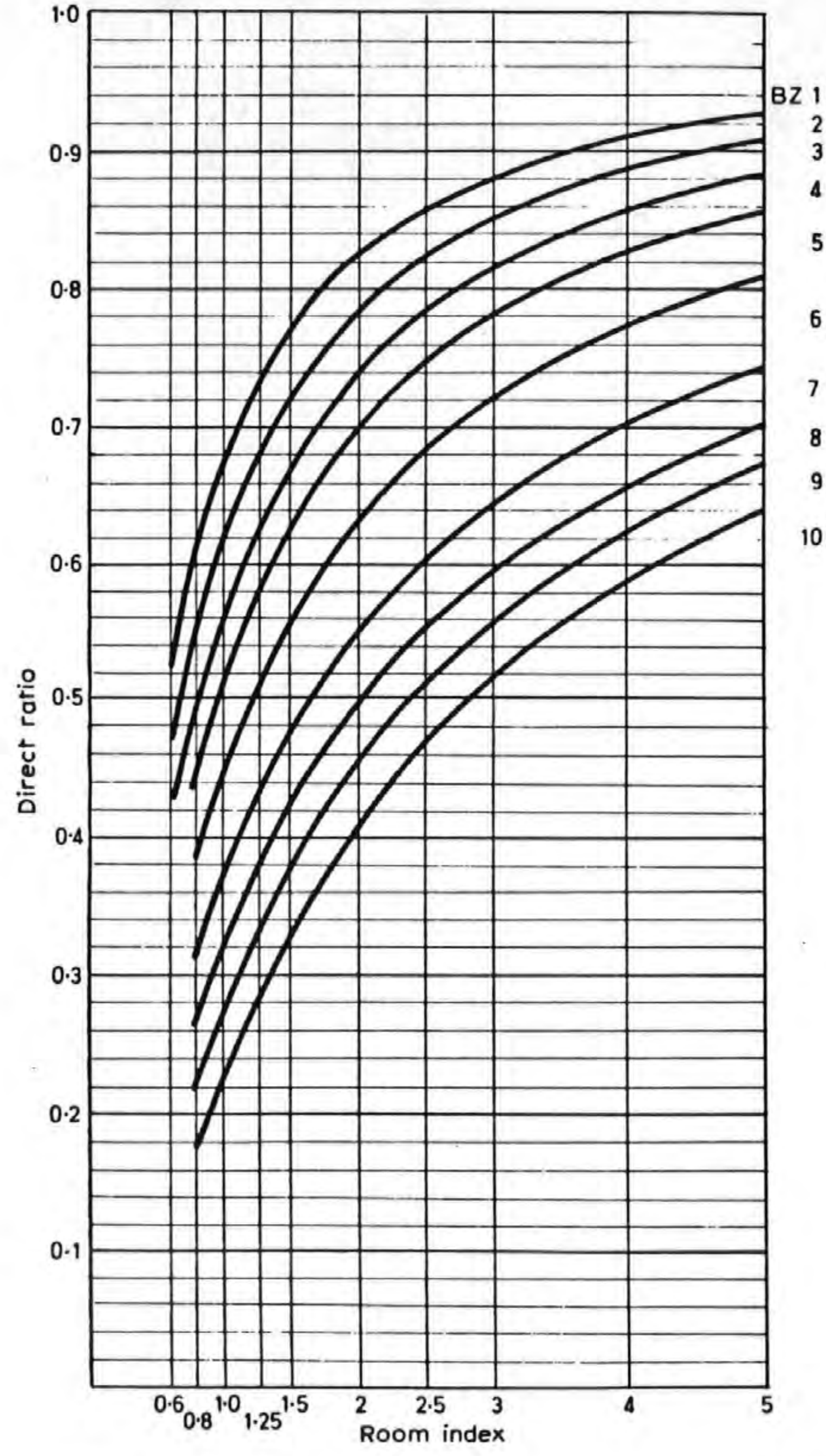


Fig. 2.05 Direct ratio as a function of room index for various BZ classifications.

In order to simplify the data input procedure for the computer model it is convenient to express the BZ curves in polynomial form as shown in Figure 2.06. Hence specification of the lighting system is made simply by inputting the value of the BZ classification, the computer being left to evaluate the direct ratio using the appropriate polynomial.

2.8 An alternative approach to luminance calculation.

In situations where the system is to be used interactively with a television display so that rapid subjective judgements may be made and appropriate changes decided upon, the computation time for the full inter-reflection theory may be considered too long. This is the case likely to apply when the system is used as a teaching aid.

As mentioned in section 2.1, the inter-reflection problem may be solved approximately, without the need to solve simultaneous equations. In 1973, Cuttle[6] developed a hybrid steady-state/iterative approach, also known as a split-flux method, which gives values for the surface luminances by direct substitution. Although in certain extreme cases, this method can lead to large inaccuracies, for most lighting situations luminance values of reasonable practical accuracy are claimed.

Using this method, the computation time for 18 luminances is of the order of 2½ seconds using the PDP 8F and FOCAL (see footnote) and thus the system becomes highly interactive.

FOOTNOTE:
'Programming Languages'; Digital Equipment Corporation; PDP8 Handbook series.

Fifth order polynomial curve fit for relationship between direct ratio Y and room index X

Direct ratio - room index taken from IES. Monograph 10 (Coomber and Jay)

$$Y = A_o + A_1X + A_2X^2 + A_3X^3 + A_4X^4 + A_5X^5$$

BZ1

$A_o = +9.309\ 652\ 890 \times 10^{-2}$
$A_1 = +1.117\ 152\ 400$
$A_2 = -7.080\ 257\ 870 \times 10^{-1}$
$A_3 = +2.361\ 763\ 620 \times 10^{-1}$
$A_4 = -3.911\ 208\ 790 \times 10^{-2}$
$A_5 = +2.521\ 230\ 470 \times 10^{-3}$

BZ2

$A_o = +7.811\ 337\ 130 \times 10^{-4}$
$A_1 = +1.148\ 316\ 950$
$A_2 = -7.001\ 586\ 610 \times 10^{-1}$
$A_3 = +2.282\ 339\ 050 \times 10^{-1}$
$A_4 = -3.724\ 951\ 620 \times 10^{-2}$
$A_5 = +2.378\ 031\ 670 \times 10^{-3}$

BZ3

$A_o = +2.602\ 553\ 760 \times 10^{-2}$
$A_1 = +9.500\ 772\ 380 \times 10^{-1}$
$A_2 = -5.169\ 233\ 020 \times 10^{-1}$
$A_3 = +1.552\ 747\ 270 \times 10^{-1}$
$A_4 = -2.388\ 559\ 190 \times 10^{-2}$
$A_5 = +1.460\ 427\ 410 \times 10^{-3}$

BZ4

$A_o = -2.752\ 811\ 170 \times 10^{-2}$
$A_1 = +9.247\ 586\ 250 \times 10^{-1}$
$A_2 = -4.837\ 453\ 690 \times 10^{-1}$
$A_3 = +1.425\ 489\ 860 \times 10^{-1}$
$A_4 = -2.173\ 320\ 330 \times 10^{-2}$
$A_5 = +1.324\ 161\ 080 \times 10^{-3}$

BZ5

$A_o = +8.103\ 713\ 480 \times 10^{-4}$
$A_1 = +7.702\ 875\ 210 \times 10^{-1}$
$A_2 = -3.621\ 824\ 080 \times 10^{-1}$
$A_3 = +9.821\ 225\ 770 \times 10^{-2}$
$A_4 = -1.399\ 326\ 450 \times 10^{-2}$
$A_5 = +8.062\ 637\ 780 \times 10^{-4}$

BZ6

$A_o = -2.638\ 627\ 590 \times 10^{-2}$
$A_1 = +6.369\ 903\ 230 \times 10^{-1}$
$A_2 = -2.773\ 253\ 400 \times 10^{-1}$
$A_3 = +7.322\ 725\ 800 \times 10^{-2}$
$A_4 = -1.039\ 867\ 830 \times 10^{-2}$
$A_5 = +6.036\ 283\ 820 \times 10^{-4}$

BZ7

$A_o = -3.739\ 417\ 700 \times 10^{-2}$
$A_1 = +5.426\ 190\ 760 \times 10^{-1}$
$A_2 = -2.137\ 850\ 580 \times 10^{-1}$
$A_3 = +5.317\ 693\ 510 \times 10^{-2}$
$A_4 = -7.290\ 833\ 080 \times 10^{-3}$
$A_5 = +4.145\ 880\ 680 \times 10^{-4}$

BZ8

$A_o = -5.181\ 104\ 150 \times 10^{-2}$
$A_1 = +4.927\ 618\ 540 \times 10^{-1}$
$A_2 = -1.826\ 743\ 170 \times 10^{-1}$
$A_3 = +4.343\ 538\ 640 \times 10^{-2}$
$A_4 = -5.791\ 033\ 630 \times 10^{-3}$
$A_5 = +3.244\ 255\ 530 \times 10^{-4}$

BZ9

$A_o = -1.245\ 357\ 290 \times 10^{-1}$
$A_1 = +5.714\ 994\ 200 \times 10^{-1}$
$A_2 = -2.533\ 388\ 190 \times 10^{-1}$
$A_3 = +7.376\ 156\ 130 \times 10^{-2}$
$A_4 = -1.162\ 934\ 110 \times 10^{-2}$
$A_5 = +7.374\ 941\ 190 \times 10^{-4}$

BZ10

$A_o = -1.149\ 000\ 710 \times 10^{-1}$
$A_1 = +3.941\ 572\ 810 \times 10^{-1}$
$A_2 = -9.368\ 875\ 630 \times 10^{-2}$
$A_3 = +1.081\ 000\ 910 \times 10^{-2}$
$A_4 = -1.617\ 990\ 340 \times 10^{-4}$
$A_5 = -4.601\ 455\ 500 \times 10^{-5}$

Figure 2.06 BZ reference curves in polynomial form[19]

Cuttle uses the concept of conservation of energy after the first reflection has taken place at all the surfaces.

He derives 3 expressions for the direct illuminance of 3 regions

1. The luminaire plane $(E_d)_c$

2. The walls $(E_d)_w$

3. The reference or work plane $(E_d)_f$

These expressions are

$$(E_d)_c = \frac{\phi_{Dn}}{A_2} \times FFR \qquad 2.6$$

where ϕ_{Dn} is total downward flux and A_2 is ceiling area

$$(E_d)_w = \frac{\phi_{Dn}}{A_3} \times (1-DR)$$

$$= \frac{\phi_{Dn}}{A_2} \times k_r \times (1-DR) \qquad 2.7$$

$$(E_d)_f = \frac{\phi_{Dn}}{A_2} \times DR \qquad 2.8$$

By a knowledge of the reflection factors of the 3 areas we may find the total first-reflected flux from all the surfaces, where FFR and DR are defined in section 2.7. This flux is given by the expression:

$$\phi_{Dn} \left[(FFR \times \rho_1) + (1 - DR)\rho_3 + (DR \times \rho_2) \right]^* \qquad 2.9$$

By energy conservation, this should equal the total indirect flux absorbed by surfaces. Here an approximation is introduced. To find the total flux absorbed the following expression is used:

* ρ_1, ρ_2, ρ_3, are respectively the ceiling, floor and wall reflectances.

$$\text{flux absorbed} = E_i \times A \times (1 - \rho) \qquad 2.10$$

where E_i = average value of the indirect illuminance on to the 3 areas.

A = total surface area

ρ = average reflectance of all surfaces

The model assumes that the inter-reflected illuminance is the same for each surface.

The practical implications of this assumption are discussed later in section 2.8; under most conditions of nearly uniform lighting this is a valid approximation.

The result of the approximation made by Cuttle is to simplify greatly the derivation of the total luminances of the various surfaces. It is necessary to substitute directly into simple formulae, and solution of simultaneous equations is not required.

From equation 2.8 we write the direct illuminance of the work surface as

$$(E_d)_f = \frac{\phi_{Dn} \cdot DR}{A_2}$$

Cuttle introduces a new variable, the 'Indirect Ratio' IR where

$$IR = \frac{E_i \times A_2}{\phi_{Dn}}$$

$$= \frac{\text{total indirect flux passing through surface 2}}{\text{total downward flux}} \qquad 2.11$$

Therefore, we may use IR in 3 expressions for the total illuminance of the surfaces:

$$(E)_c = \frac{\phi_{Dn}}{A_2} (FFR + IR) \qquad 2.12$$

$$(E)_f = \frac{\phi_{Dn}}{A_2} (DR + IR) \qquad 2.13$$

$$(E)_w = \frac{\phi_{Dn}}{A_2} \times \frac{k_r}{2} \times (1 - DR + \frac{2IR}{k_r}) \qquad 2.14$$

we may write

$$= \frac{A_1\rho_1 + A_2\rho_2 + A_3\rho_3}{A_1 + A_2 + A_3}$$

$$= \frac{\rho_1 + \rho_2 + \frac{2}{k_r} \times \rho_3}{2 + \frac{2}{k_r}}$$

and $A = A_1 + A_2 + A_3 = A_2(2 + \frac{2}{k_r})$

Therefore combining 2.9, 2.10 and 2.11 we obtain

$$IR = \frac{DR \times \rho_2 + (1 - DR)\rho_3 + (FFR \times \rho_1)}{2 + \frac{2}{k_r}(1 - \rho_3) - \rho_1 - \rho_2}$$

thus IR may be calculated directly.

For many practical lighting situations the above approximation has minimal influence on the accuracy of the final result. It is clear, however, that under certain circumstances the method can give rise to gross errors.

As would be expected, the accuracy of the method decreases as the indirect component of illuminance increases. From the definitions of DR and IR we see that for totally indirect lighting FFR = ∞ and IR = ∞, hence the error would be infinite. However, for many practical situations where FFR $\leq$ 1.0 the accuracy will generally be within 10% and for FFR = 0 the results are correct to within 3%.

To summarise, the split-flux method, being based on a simplified model, is less accurate than methods employing full inter-reflection theory. In cases dominated by indirect lighting the errors can become unacceptably large. However, the method has the advantage of providing extremely rapid results since far less arithmetic is involved. Thus, where accuracy is not the primary requirement, for example where the effects of certain lighting changes are being demonstrated, in a teaching situation, the Cuttle technique provides a rapid means of lighting prediction which may be repeated with new parameters with a minimum delay.

In the context of the computer controlled colour display system described in this thesis, the Cuttle method gives an almost instantaneous response from the computer allowing a student for example, to see without delay the influence of the various parameters involved in the lighting calculation. In cases where accuracy is an over-riding requirement, the full inter-reflection model may be employed with the consequent increase in calculation time.

References

1. Harrison, W. and Anderson, E. A.
   Trans. Illum. Eng. Soc. (New York), vol. 16, p. 97 (1920).

2. Stokes, G.
   Proc. Roy. Soc., vol. 11, p. 545 (1862).

3. Sumpner, W. E.
   Proc. Phys. Soc., vol. 12, p. 10 (1892).

4. Ulbricht, R.
   Das Kugelphotometer, Oldenbourg, Munich, 1920.

5. Phillips, R. O.
   Trans. Illum. Eng. Soc. (London), vol. 21, p. 75 (1956).

6. Cuttle, C.
   Light and Lighting, vol. 66, p. 349 (1973).

7. Moon, P. and Spencer, D. E.
   J. of Franklin Institute, vol. 242, p. 111 (1946).

8. Yamauti, Z.
   J. of Opt. Soc. Amer., vol. 13, no. 5, p. 561 (1926).

9. Waldram, J. M.
   Trans. Illum. Eng. Soc. (London), vol. 19, p. 95 (1954).

10. Dourgnon, J.
    Cahiers de Centre Scientifique et Technique du Batiment, no. 27, cahier 229.

11. O'Brien, P. F. and Howard, J. A.
    Illum. Engng., vol. 54, p. 209 (1959).

12. O'Brien, P. F.
    Proc. C.I.E., Section 3.1.1, Brussels, (1959).

13. Lynes, J. A.
    I.E.S. Monograph no. 1, Illuminating Engineering Society (London) 1959.

14. I.E.S. Technical Report no. 2, 2nd edition.
    Illuminating Engineering Society (London) 1971.

15. Moon, P.
    The Scientific Basis of Illuminating Engineering. p. 328, McGraw-Hill, New York, 1936.

16. Buckley, H.
    Phil. Mag., vol. 4, p. 753 (1927).

17. Hamilton, D. C. and Morgan, W. R.
Radiant-Interchange Configuration Factors. National Advisory Committee for Aeronautics, Technical Note 2836 (1952).

18. Spencer, D. E.
Illum. Engng., vol. 53, p. 243 (1958).

19. Simpson, J. W.
I.E.S. Symposium. The use of computers for lighting calculation. University College London, March 30th (1971) Report published March 1972.

Chapter 3

A COMPUTER CONTROLLED COLOUR TELEVISION DISPLAY SYSTEM AN AID FOR THE LIGHTING DESIGNER.

3.1 Introduction.

In chapter 2, methods were described for the prediction of luminance distributions in lighted interiors. The techniques, although straightforward, are too laborious for hand calculation but lend themselves to solution by digital computer. Results obtained in this way may be readily interpreted by an experienced lighting designer, but are likely to be of little value to the novice, particularly in a non-numerate discipline. Experience in the Plymouth Polytechnic School of Architecture served to emphasize this latter problem, leading to the development of the display system to be described in this chapter. The system uses a digital computer to provide both rapid results and an interactive dialogue with the student, combined with a television simulation of the room interior to provide the results in an immediate visual form.

Lighting designers are concerned about the quality as well as quantity of light in a particular environment. The quality of incident light is a function of its direction and its colour rendering properties. The colour rendering performance of a light source may be calculated, for a surface viewed in isolation, simply from the spectral power distribution of the source and spectral reflectance characteristics of the surface colour, a procedure described in section 3.3. An approach to the solution of the more general problem, in which a surface colour is viewed under direct and indirect illumination involving inter-reflection with the other coloured surfaces, has been suggested by Lynes (1957)[1]. The method involves subdividing the spectral power distribution of the light source into a number of narrow bands, for each of which the surface reflectance and illumination are considered invariant with wavelength. In this way the designer may take account of the influence of both the light source and inter-reflection on surface chromaticity.

Taking advantage of the wide colour-reproducing capability of the television display, the system was designed to take account of surface chromaticity and reflectance in addition to lamp colour rendering.

This chapter describes the systems used for quantifying colour, with a discussion of additive colour mixture having particular reference to colour television. Techniques are discussed for colour and luminance measurement, essential preliminaries if accurate colorimetric reproduction is to be achieved. The reader familiar with colorimetric techniques may therefore prefer to omit section 3.3. A general method is then given for solution of the inter-reflection problem taking account of the power spectrum of the light source and reflectance characteristics of the surfaces. Some disadvantages of this approach are discussed along with alternative methods which may be preferred, depending on the limitations of the computer.

3.2 System specification.

The display produces a simple perspective simulation of a rectangular room interior. The viewing position may be varied under computer control.

Data is input to define the following:

1. The room dimensions.
2. The number of light fittings.
3. The type of light fitting.
4. Total power input.
5. The type of lamp.
6. The colour and reflectance specification of the room surfaces.

The software then takes account of inter-reflection in calculating the luminance and chromaticity of each surface. Account is taken of the chromaticity as well as reflectance of each surface in the calculation. Briefly the display specification is:

1. The system has a chromaticity resolution better than 0.001 units in CIE uniform colour space over most of the luminance range.

2. Colours may be defined using the CIE 1931 x,y chromaticity diagram, the CIE 1960 UCS diagram, or the Munsell Renotation System.

3. The direct light distribution from the luminaires is simply described in terms of British Zonal (BZ) number and flux fraction ratio (FFR).

4. The colour rendering properties of the light source are included in the calculation.

5. Simplified inter-reflection routines are available for computers with limited storage capacity or slow response.

6. The display is compatible with the standard CCIR 625 line system 1 television format and may therefore be used with a standard studio colour monitor.

7. Communication with the computer is provided by an RS232 or a 20 mA current loop interface with variable baud rate. Thus the system is compatible with virtually any digital computer from microprocessor to mainframe.

8. The picture is generated by hardware, but avoids the use of a frame store to minimise cost and provide rapid data transfer.

3.3 Quantifying colour.

If a group of people with normal colour vision is asked to judge the closeness of matching between a colour sample and some standard reference, a wide variety of descriptions of the colour difference is likely to result. Such subjective disagreement can lead to embarrassment, particularly in a commercial environment.

Attempting to introduce a more quantitative approach to colour judgement, the International Commission on Illumination (C.I.E.) in 1931 recommended a method of colour measurement based on a 'standard observer'. The 'standard observer' was defined using results of colour matching experiments carried out by Guild at the National Physical Laboratory (1931)[2] on seven observers, combined with similar measurements by Wright at the University of London (1928-1929)[3] using ten observers.

The matching experiment used a bipartite field as illustrated in Figure 3.01. One side of the field was illuminated by pure spectral stimuli produced by a monochromator. The other side of the field could be adjusted to match by varying the mixture of three 'primary' lights. Since no mixture of three lights can match all spectral colours, at some wavelengths it is necessary to have negative amounts of colour stimulus from the matching field. In practice, the match is achieved by adding small quantities of the 'negative' primary to the spectral field. In order to avoid colour matching curves having negative quantities at certain wavelengths, the C.I.E. used the expedient of imaginary super-saturated primary stimuli X, Y and Z. The colour matching functions $r(\lambda)$, $g(\lambda)$ and $b(\lambda)$ obtained by experiment (see Figure 3.02) may be transformed into the quantities $x(\lambda)$, $y(\lambda)$ and $z(\lambda)$, the amounts of the imaginary primaries required to match any spectral stimulus. As shown in Figure 3.03 all real colours are enclosed by a triangle with apices X, Y and Z. The colour matching functions of Figure 3.04 therefore have all positive values. The 1931 colour matching functions are derived for a stimulus subtending 2° at the eye; further data for 10° field were published in 1964.

The quantities $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ are those required to match pure spectral line frequencies. Most visual stimuli present the viewer with relatively broad spectral distributions. An arbitrary

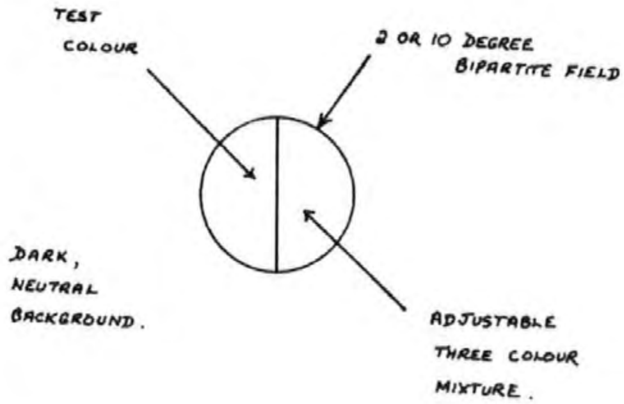


Fig. 3.01 The colour matching experiment.

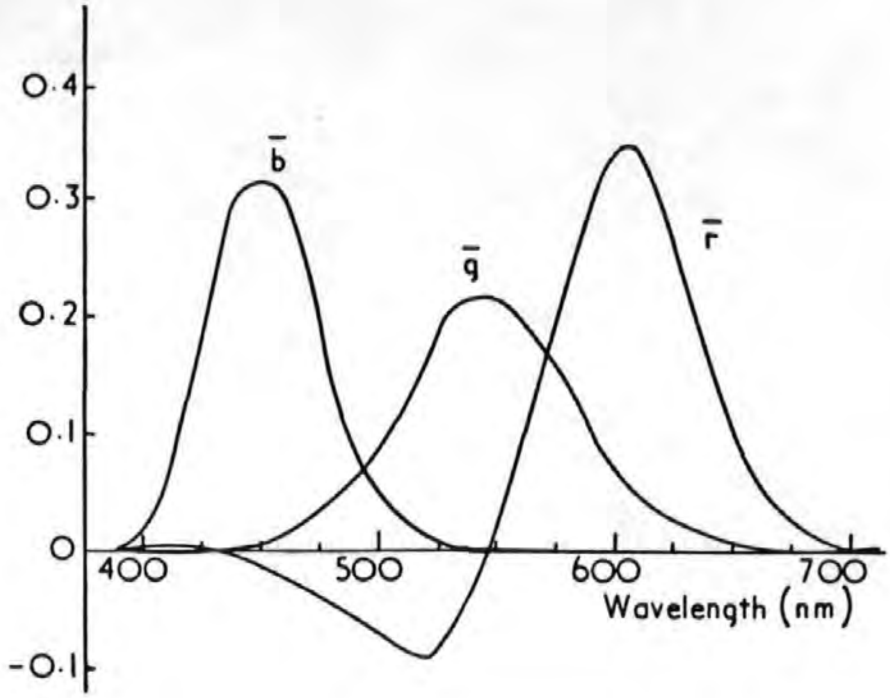


Fig. 3.02 The $\overline{r}$ $\overline{g}$ $\overline{b}$ colour matching functions

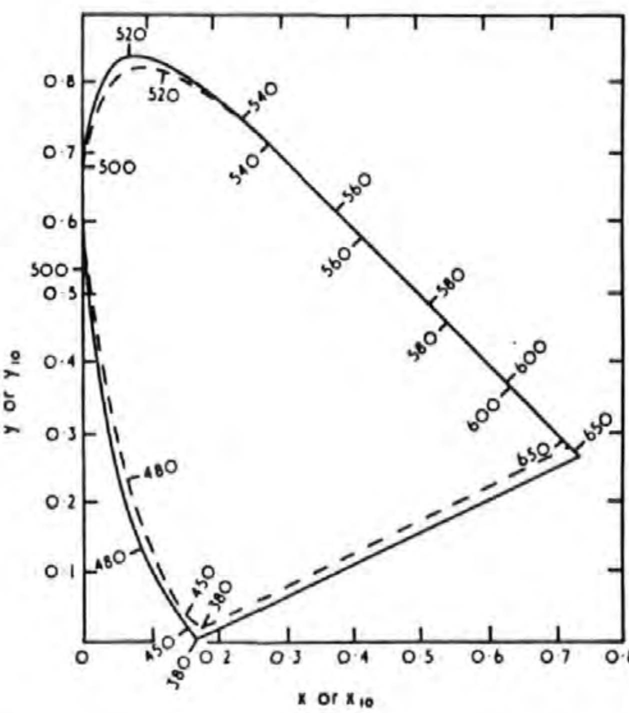


Fig. 3.03 The C.I.E. x y chromaticity diagram.

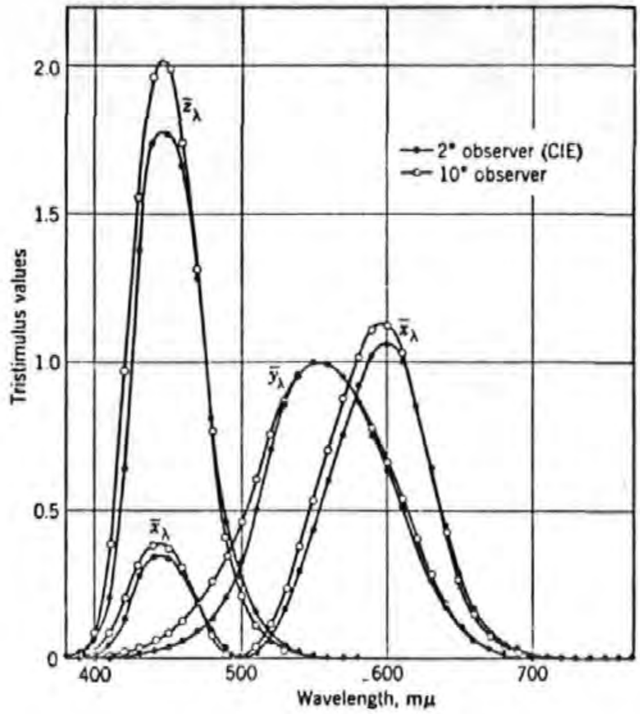


Fig. 3.04 The C.I.E. all positive colour matching functions.

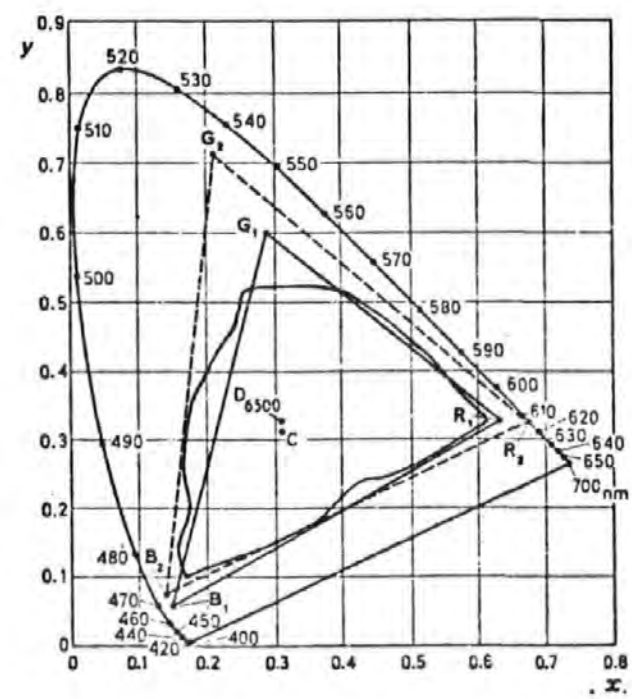


Fig. 3.05 The N.T.S.C (2) and PAL (1) chromaticity gamuts.

distribution of luminance $L(\lambda)$ may be regarded as an infinite number of spectral components where

$$L(\lambda) = E(\lambda)\,\rho(\lambda)$$

with $E(\lambda)$ the illuminance distribution and $\rho(\lambda)$ the reflectance distribution.

The 'tristimulus' values X, Y and Z required to match a given surface colour may now be calculated using

$$X = k \int_0^\infty E(\lambda)\rho(\lambda)\bar{x}(\lambda)\,d\lambda$$

$$Y = k \int_0^\infty E(\lambda)\rho(\lambda)\bar{y}(\lambda)\,d\lambda$$

$$Z = k \int_0^\infty E(\lambda)\rho(\lambda)\bar{z}(\lambda)\,d\lambda$$

The colour matching functions allow specification of a colour stimulus by using the objectively measured spectral distribution of radiant flux from that stimulus. The final specification is reduced from a continuous function to the three numbers X, Y and Z. In addition to allowing representation of all physical colour stimuli with positive co-ordinates, the positions of the X, Y, Z primaries in colour space were selected such that all the luminance information is contained in the value of Y. This convenience results from the $\bar{y}(\lambda)$ colour matching function being made identical with the $V(\lambda)$ relative luminosity function for the standard observer.

3.31 The C.I.E. 1931 chromaticity diagram.

Colours may be specified in a convenient graphical form by normalizing the tristimulus values X, Y and Z to produce 'chromaticity' co-ordinates x, y and z using the equations:-

$$x = \frac{X}{X + Y + Z}$$

$$y = \frac{Y}{X + Y + Z}$$

$$z = \frac{Z}{X + Y + Z}$$

where the new quantities are the proportions of the X, Y and Z primaries respectively in a given mixture. Since it is not possible to produce colours of a given hue with more saturation than pure spectral stimuli, a graphical representation of chromaticity will have the spectrum locus as its outer boundary. Since the C.I.E. colour matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ were derived for line spectra, the chromaticity of any point on the spectrum locus is defined by

$$x = \frac{\bar{x}(\lambda)}{\bar{x}(\lambda) + \bar{y}(\lambda) + \bar{z}(\lambda)} \quad \text{etc.}$$

The 1931 C.I.E. chromaticity diagram is a two dimensional plot of co-ordinates x,y as illustrated by Figure 3.03. The co-ordinates x,y are a measure of the dominant wavelength and purity, but do not themselves provide a complete objective description of a colour stimulus. A third dimension, perpendicular to the x,y plane, is required to define the luminance.

Figure 3.05 shows the chromaticities of the red, green and blue colour television phosphors as used for the PAL system 1 and NTSC systems. The chromaticity of any mixture of two coloured lights such as the television phosphors, lies on the straight line joining them on the chromaticity diagram. Furthermore, the chromaticity of a mixture of three primaries is limited to the area within a triangle having as apices the three phosphor chromaticities. Thus the triangles shown in Figure 3.05 indicate the range of chromaticities which may be reproduced by the two systems. A further discussion of additive colour mixture is given in section 3.34.

3.32 The 1960 C.I.E. uniform colour space.

If visual stimuli having approximately equal perceived colour differences, such as those contained in the Munsell Book of Color, have their chromaticities plotted in the 1931 C.I.E. x,y chromaticity diagram, the measured spacing is likely to be anything but equal. This non-uniformity, illustrated in Figure 3.06 in which all the lines represent equal perceived colour differences, may be eliminated by a non-linear transformation. However, a good approximation to uniformity can be achieved using the linear transformation proposed by MacAdam (1937)[4]. The new stimuli U, V, W are related to X, Y, Z:-

$$U = 2/3\ X$$

$$V = Y$$

$$W = 1/2\ X + 3/2\ Y + 1/2\ Z$$

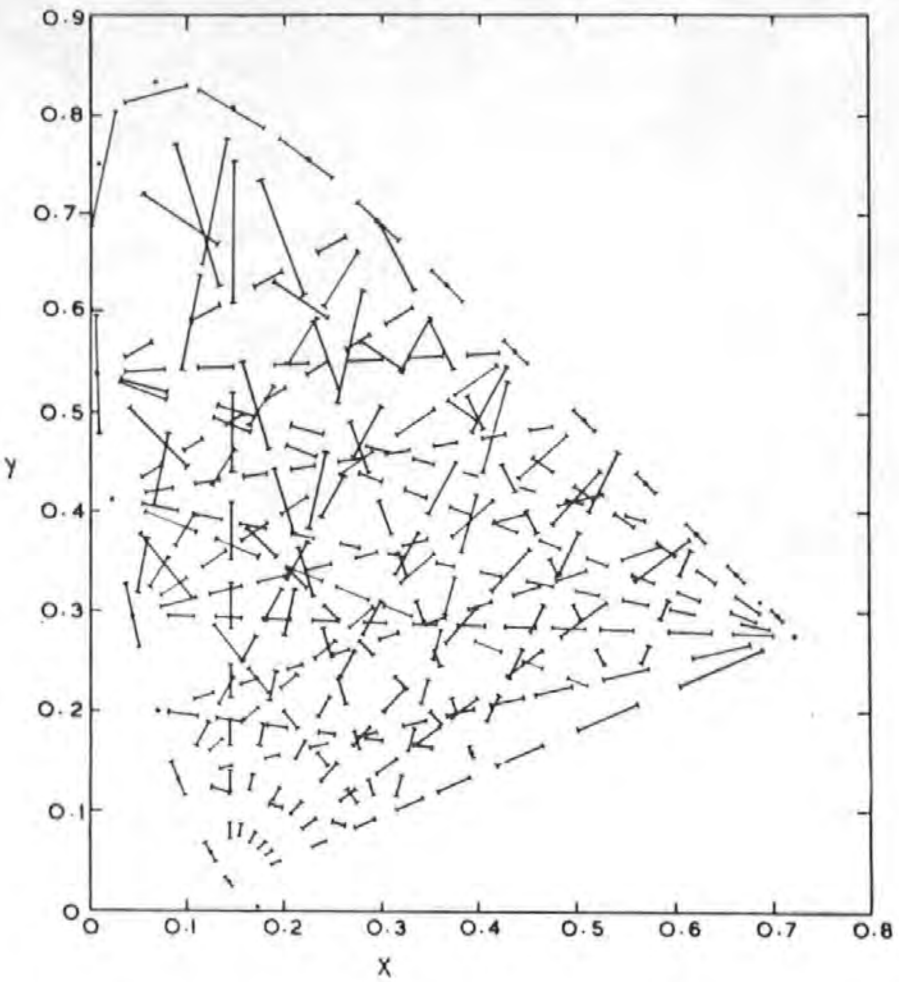


Fig. 3.06 Equal perceptual colour differences in the C.I.E. x, y chromaticity diagram.

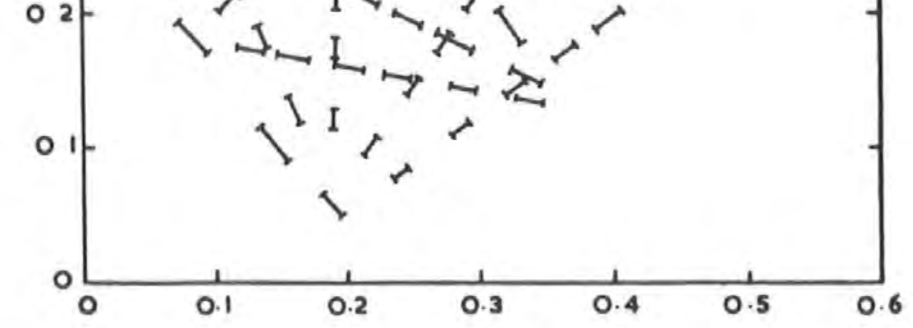


Fig. 3.07 Equal perceptual colour differences in the C.I.E. UCS diagram.

The new colour system retains a useful feature of the X,Y,Z system in that one tristimulus value, V, is made proportional to luminance.

The C.I.E. uniform colour diagram based on U,V, W has co-ordinates u,v related to x,y by

$$u = \frac{4x}{-2x + 12y + 3}$$

$$v = \frac{6y}{-2x + 12y + 3}$$

The closeness to perceptual uniformity achieved by this modified colour space is illustrated by Figure 3.07.

3.33 Modified uniform colour space.

Although significantly more uniform than the C.I.E. 1931 X,Y,Z colour system, the C.I.E. 1960 uniform colour space still fails to show equal differences in perceived colour as precisely equal distances in the u,v diagram. Figure 3.08 shows loci of constant Munsell hue and saturation plotted on a U.C.S. chromaticity diagram. If, as is claimed, the colour increments of the Munsell system are perceptually equal, the loci of constant chroma should plot as circles in a perfectly uniform C.I.E. chromaticity diagram. Clearly this is not so, as loci of constant chroma produce approximately elliptical contours.

In an attempt to improve the C.I.E. 1964 'uniform' colour space, a C.I.E. committee in 1974[5] suggested a modified system in which the v co-ordinate of the u,v chromaticity diagram becomes v' where v' = 1.5 v.

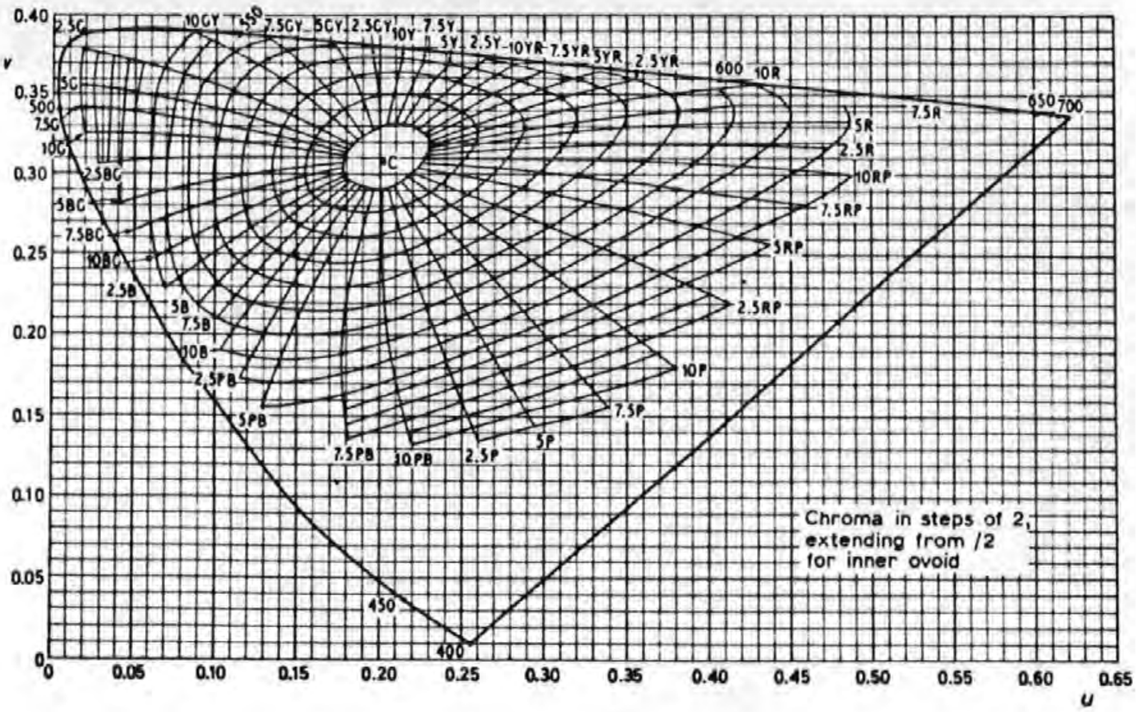


Fig. 3.08(a) Loci of constant Munsell renotation hue and chroma (value = 4) in the C.I.E. UCS diagram.

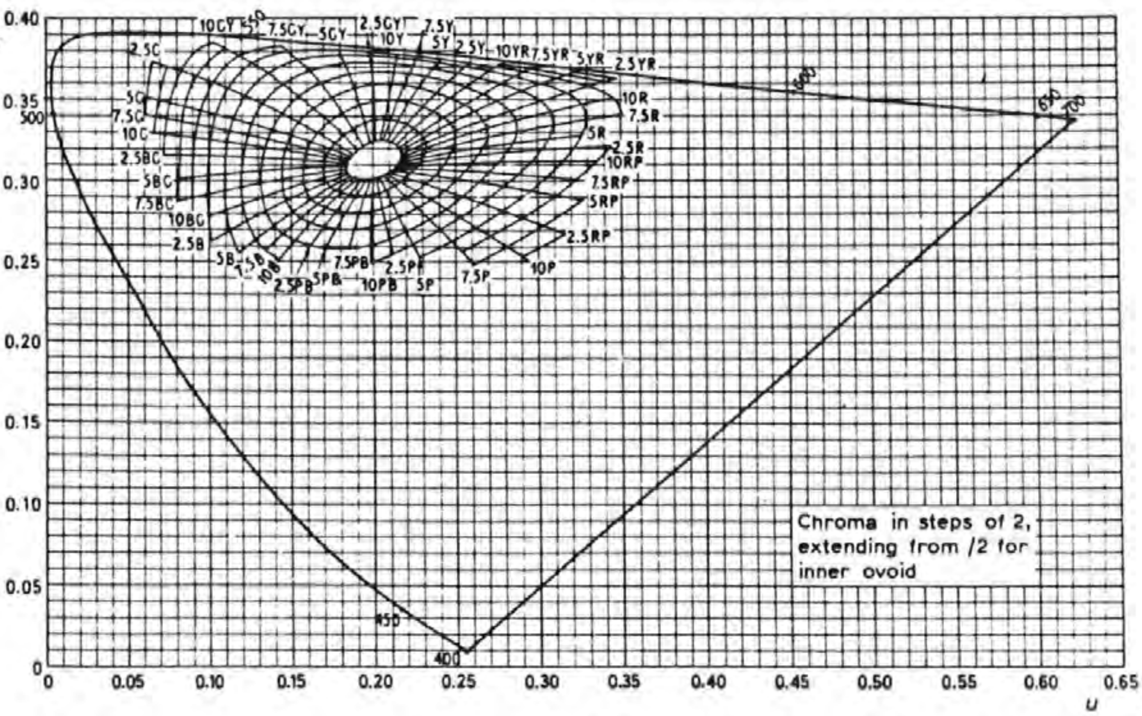


Fig. 3.08(b) Loci of constant Munsell renotation hue and chroma (value = 7) in the C.I.E. UCS diagram.

Thus $u' = u = \frac{4x}{-2x + 12y + 3}$

$$v' = 1.5v = \frac{9y}{-2x + 12y + 3}$$

For colour differences typical of those found in the Munsell Book of Color, this modification is claimed to give closer approach to perceived uniformity in the chromaticity diagram. The contours shown in Figure 3.08 appear to support this view. For this reason the work by Lynes (1974)[6] and the work discussed in section 3.65 is based on use of this later system.

3.34 Colour mixture in C.I.E. uniform colour space.

If two lights with respective luminances L_1, L_2 and chromaticities (u_1, v_1), (u_2, v_2) are added, the resulting mixture will have a chromaticity and luminance derived by the following procedure.

The sources u_1, v_1, L_1 and u_2, v_2, L_2 have tristimulus values U_1, V_1, W_1 and U_2, V_2, W_2 where

$$u_1 = \frac{U_1}{U_1 + V_1 + W_1} = \frac{U_1}{S_1} \qquad u_2 = \frac{U_2}{U_2 + V_2 + W_2} = \frac{U_2}{S_2}$$

Since in the C.I.E. system the V tristimulus value is proportional to the luminance

$$\therefore \quad v_1 = \frac{L_1}{S_1} \qquad v_2 = \frac{L_2}{S_2}$$

$$\text{for the mixture} \quad u_M = \frac{U_M}{S_M} = \frac{U_1 + U_2}{S_1 + S_2}$$

$$= \frac{u_1 S_1 + u_2 S_2}{S_1 + S_2}$$

$$= \frac{u_1 L_1/V_1 + u_2 L_2/V_2}{L_1/V_1 + L_2/V_2}$$

$$\text{and} \quad v_M = \frac{V_M}{S_M} = \frac{L_1 + L_2}{S_1 + S_2}$$

$$= \frac{L_1 + L_2}{L_1/v_1 + L_2/v_2}$$

Thus the chromaticity co-ordinates of the mixture are a weighted average of those for the constituent light sources. The weights p_1, p_2 are equal to L_1/v_1 and L_2/v_2 respectively.

The co-ordinates of the mixture are therefore

$$u_M = \frac{u_1 p_1 + u_2 p_2}{p_1 + p_2}$$

$$v_M = \frac{v_1 p_1 + v_2 p_2}{p_1 + p_2}$$

As the ratio L_1/L_2 is changed, the locus of the chromaticity of the mixture forms a straight line joining u_1, v_1 and u_2, v_2. Similarly for a mixture of three lights L_1, L_2 and L_3

$$u_M = \frac{u_1 p_1 + u_2 p_2 + u_3 p_3}{p_1 + p_2 + p_3}$$

$$v_M = \frac{v_1 p_1 + v_2 p_2 + v_3 p_3}{p_1 + p_2 + p_3}$$

In this case the chromaticity $u_M v_M$ is fixed by the ratio L_1 : L_2 : L_3 and confined to a triangular area in colour space having apices $(u_1 v_1)$, $(u_2 v_2)$ and $(u_3 v_3)$.

3.35 Colour-order systems.

The C.I.E. colour systems described in 3.3 are based on colour matches achieved by a 'standard observer' to a series of spectral stimuli. Objective measurements may be made which give the tristimulus values X,Y,Z or U,V,W. For the standard observer any colour stimuli producing the same tristimulus values will give the same colour sensation, regardless of its spectral power distribution. In many practical viewing situations such measurement is at best inconvenient and probably impossible. In this case it is usual to quantify object colours by reference to a set of material standards.

To meet the demand for a convenient reference source of object colours, several attempts have been made to produce material standards whose colour attributes form a rational order. Three approaches to this problem have been taken, systems being based on additive colour mixture, such as the Ostwald system, subtractive colour mixture such as the Plochere system and colour perception as used to develop the Munsell and DIN colour systems.

The systems based on colour mixture have samples which are spaced non-uniformly through the colour solid. These systems do not directly represent the perceptual attributes of colour. These shortcomings have led to the development of systems based not on objective colour mixture, but on colour appearance.

The Munsell colour system is based on a set of material standards for which three perceptual attributes are judged by an observer with normal colour vision, under specific viewing and illumination conditions. The three attributes are Hue, Value and Chroma. Hue is that quality of an object colour associated with its dominant wavelength; it is this which distinguishes, for example, a violet from a green. Value is a measure of the reflectance of the colour under the specified light source. Chroma is related to the saturation of the object colour. Each variable claims to represent equal perceived increments. Hues are arranged circularly about a central axis. Distance along the axis represents value and distance away from the central axis represents chroma, so as to form a colour solid (see Figures 3.09 and 3.10). The system is summarized in the Munsell Book of Color[7] first published in 1929.

3.36 The Munsell Renotation System.

In 1943, a sub-committee organised by the Optical Society of America suggested modification of the original Munsell Book of Color (Newhall, Nickerson and Judd, 1943)[8]. These modifications were based on numerous visual estimates of the spacing of the samples contained in the Book of Color and are thought to approach more closely

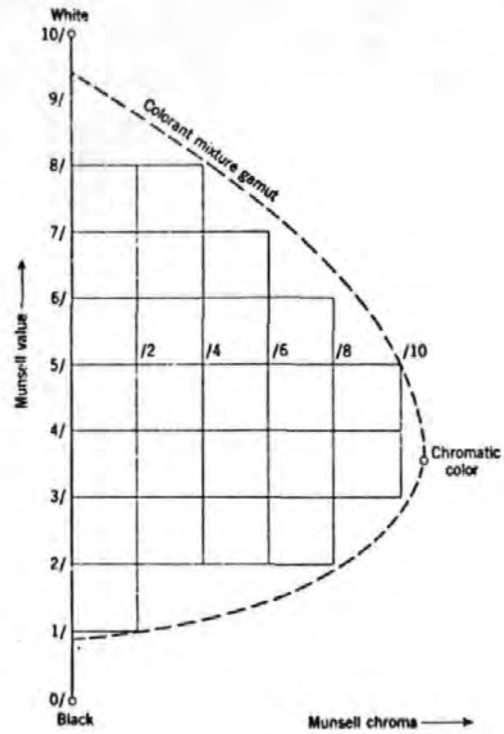


Fig. 3.09 Munsell chroma gamut.

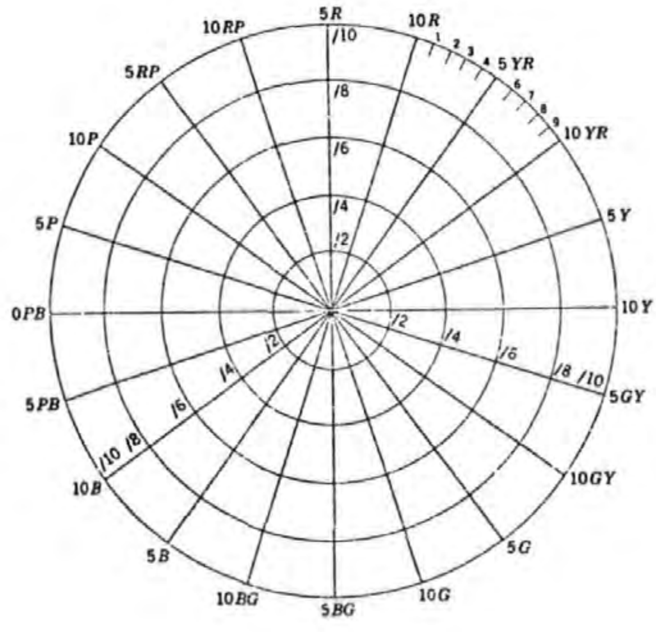


Fig. 3.10 Munsell hue gamut.

the original aim of providing samples of practical usefulness and psychological equispacing. On the basis of this revised spacing, tables were published in terms of the C.I.E. 1931 colour space, the revised system being known as the Munsell Renotation. Figure 3.08(a) shows the Munsell Renotation for value 4 transformed to the 1964 C.I.E. uniform colour space. The popularity of the Munsell system amongst research workers has led to the publication of conversion charts and tables allowing the expression of Munsell Renotation directly in the C.I.E. system of units and vice versa. It is possible to convert from the Munsell value V_M scale to V using the definition

$$V = 1.2219\ V_M - 0.12111\ V_M^2 + 0.23951\ V_M^3 - 0.021009\ V_M^4 + 0.0008404\ V_M^5$$

For a perfect white $V_M = 10$ and $V = 102.568$

$$\therefore \text{ the reflectance } \rho = V/102.568$$

Similar direct conversions for hue and chroma are unfortunately not possible although these have been generated by digital computer (Keegan et al, 1958[9]; Rheinboldt and Menard, 1960[10]).

The Munsell Renotation System, therefore, gives a convenient material reference for colour specifications given in terms of C.I.E. chromaticity and has been used as a source of reference colours for this work.

3.37 Three dimensional perceptually uniform colour spaces.

3.37.1 The C.I.E. U*V*W* system.

In Figure 3.08(a), chromaticities are shown for Munsell Renotation samples of value 4 plotted in the C.I.E. 1960 UCS diagram. Although far from perfect, the diagram gives an approximation to perceptual uniformity, certainly superior to the C.I.E. 1931 system. In Figure 3.08(b), the chromaticities of samples with value increased to 7 are shown, illustrating the sensitivity of colour perception to reflectance. To provide a means of quantifying perceived colour difference taking account of a third dimension, the C.I.E. have recommended the U*, V*, W* system, effectively the 1960 C.I.E. uniform colour space extended to include brightness (Wyzecki, 1963)[11]. The three quantities U*, V*, W* are defined:-

$$U^* = 13\,W^*\,(u - u_o)$$

$$V^* = 13\,W^*\,(v - v_o)$$

$$W^* = 25\,Y^{1/3} - 17$$

where u_o, v_o is the chromaticity of the colour perceived as achromatic

u, v is the chromaticity of the colour sample

Y is the relative luminance of the colour on a scale 1 to 100

W* is a measure of brightness. It is known that for a dark surround, the brightness of a display screen is approximately proportional to the cube root of its luminance (Bartleson and Breneman, 1967)[12]. In these conditions, therefore, W* will provide a satisfactory measure of the display brightness.

As discussed in section 3.23, the C.I.E. have proposed an improvement to the 1960 uniform colour space in which u' = u and v' = 1.5 v. This modification also applies to the U*, V*, W* system giving:-

$$U' = U^*$$

$$V' = 1.5\,V^*$$

$$W' = W^* + 1$$

The significance of this scaling up of the v dimension of the 1960 C.I.E. chromaticity diagram is made evident by Figure 3.08. If the Munsell system is taken as a reference for perceptual uniformity, then loci of constant Munsell chroma should map as circles in a perceptually uniform chromaticity diagram. These loci, shown in Figure 3.08, are brought closer to this ideal in the new C.I.E. chromaticity system.

3.37.2 The CIELAB system.

The C.I.E. U'V'W' colour space is designed to define colour differences which are comparable with the colour increments found in the Munsell Book of Color. In some applications of colour difference measurement, such as in the textile industry, smaller colour differences become important. For this purpose, the C.I.E. (1974)[5] proposed an alternative system based on the Adams-Nickerson

ANLAB formulae, which became known as the CIELAB system. The CIELAB colour space has three dimensions defined as:

$$L = 25(100Y/Y_o)^{1/3} - 16 \qquad 1 \leqslant Y \leqslant 100$$

$$A = 500\left[(X/X_o)^{1/3} - (Y/Y_o)^{1/3}\right]$$

$$B = 200\left[(Y/Y_o)^{1/3} - (Z/Z_o)^{1/3}\right]$$

where X_o, Y_o, Z_o define the tristimulus values of a white-appearing stimulus.

The formula for lightness index L is identical with that used for the U'V'W' system. Whereas simple linear transforms were used to relate U'V'W' co-ordinates to the C.I.E. X,Y,Z system, a similar transform for the LAB co-ordinates unfortunately does not exist.

3.4 Colour spaces and computer graphics.

Computer controlled colour display systems usually involve three distinct components, a man/machine interface, software algorithms and finally a display device. Colours are usually described in terms of a three dimensional co-ordinate system; unfortunately, no one colour system has found universal acceptance. For a colour graphics display it is likely that two or more colour spaces are needed to best suit the needs of the system.

In the case of a colour television display, coloured stimuli are matched by the additive mixture of red, green and blue light sources. The mixture (m) may be described graphically as the sum of three orthogonal vectors R, G and B. For the purposes of the picture display device, it is usually most convenient to work in R,G,B colour space. The user interacting with a computer colour display system is not likely to find R,G,B colour space either meaningful or convenient. Subjectively, colour is usually described in terms of three attributes; hue (related to dominant wavelength and probably the most immediate attribute), saturation (a measure of the departure from neutrality), and lightness (a measure of reflectance), as described in 3.4. Two commonly used colour systems, the Munsell system and the C.I.E. 1960 colour space, although developed in different ways and each having advantages for different purposes, provide, for controlled viewing conditions, some measure of these attributes. Some users may be more accustomed to the subtractive colour 'primaries' cyan, yellow and magenta, familiar to artists. For the purposes of man-machine interaction it is preferable to be able to select a colour space to suit the application.

The central area of the display system, the computation, may also benefit from a careful selection of colour space. For example, the simulation of specular reflections from coloured surfaces is achieved by a colour series starting at black, increasing saturation and luminance in the particular hue direction and decreasing saturation to end at or near white. Definition of such a colour scale in the Munsell system would be more convenient than, for example, R,G,B space.

Simulation of the reflection of coloured light by coloured objects may be achieved by multiplication of the spectral power distribution of the light source and the spectral reflectance distribution of each surface (see section 3.51). The amount of computation involved in this procedure is much reduced by an approximation in which the chromaticity of the light source is transformed into three components in, for example, R,G,B space. A similar transformation for the reflectance distribution yields three equivalent reflectances ρ_R, ρ_G and ρ_B from which the chromaticity of the reflected light is readily obtained. A discussion of this approximation and its shortcomings is given in section 3.62.

To summarise, computer colour displays will work most effectively if the appropriate colour space is selected for each element of the system. Further discussion of graphic display colour spaces is given by Booth and Schroeder (1977)[13], Joblove and Greenberg (1978)[14], Smith (1978)[15] and Sloan and Brown (1979)[16].

3.5 Television chromaticity reproduction under computer control.

3.51 The television colour gamut.

The discussion of 3.34 shows that a mixture of two coloured lights will produce a chromaticity lying on a straight line joining the two original chromaticities plotted on a C.I.E. chromaticity diagram. Further, any mixture of n coloured lights will have a chromaticity lying within the convex polygon formed by linking the n starting chromaticities.

Figures 3.11 and 3.12 show the chromaticities associated with various common objects with daylight illumination, plotted on the 1960 C.I.E. UCS diagram. Since the shape enclosed by the spectrum locus is

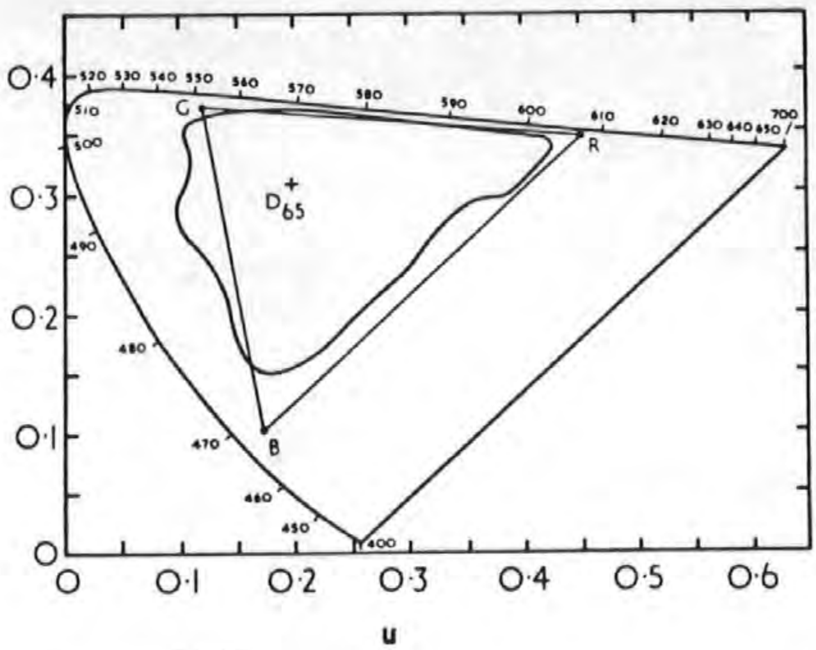


Fig. 3.11 Television colour gamut with chromaticity range of surface colours.

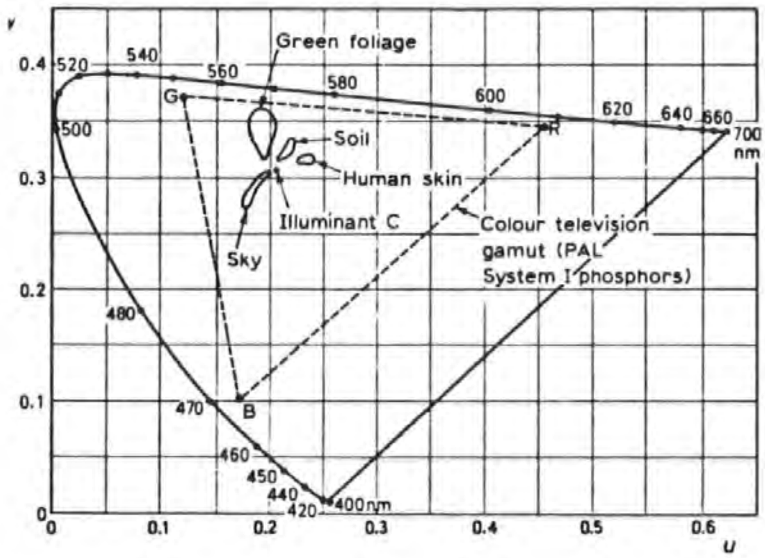


Fig. 3.12 The chromaticity range for common stimuli.

approximately triangular, a large proportion of the colour range may be reproduced by mixture of three coloured lights with chromaticities located near the red, green and blue spectral extremities of the diagram. Clearly, additional 'primary' light sources could further extend the colour gamut, but this becomes a case of diminishing returns. For this reason, three primary colours, red, green and blue, are used for television colour reproduction.

Further examination of Figures 3.11 and 3.12 indicates that most common objects have chromaticities lying within a fairly restricted central area of the diagram. This has practical advantages for colour television. Manufacture of the display tube phosphors involves a compromise between spectral purity and luminous efficiency. In the early days of American colour television the National Television Systems Committee (N.T.S.C.) recommended the use of the primary chromaticities shown in Figure 3.05, produced by silicate phosphors. In the light of experience gained with the N.T.S.C. system, the primary chromaticities chosen for the later European PAL system were chosen to sacrifice some range of chromaticity for improved luminous efficiency and hence higher screen luminance (Wood and Sproson, 1971)[17]. These primaries are also illustrated in Figure 3.05. The chromaticities falling outside the colour gamut, although representing a substantial area of the C.I.E. UCS diagram, are not representative of common colour stimuli, allowing good colour reproduction from what may initially appear to be a very restricted colour gamut.

The preference for higher screen luminance at the expense of colour saturation is consistent with the experimental results obtained by Hunt (1965)[18] and Bartleson and Witzel (1967)[19] which show that increased luminance produces a subjective increase in colour saturation. Finally, higher screen luminance serves to offset the desaturating effect of light scattered from the screen due to ambient illumination.

Having fixed the red, green and blue phosphor chromaticities, in the context of the television display, chromaticity is defined by the three luminances R, G and B. Since surface colours are normally described by their chromaticity and reflectance u_s, v_s and ρ_s under a specified light source, a routine transformation to R, G and B is required.

Transformation from one set of tristimulus values R,G,B to an equivalent set based on other primaries, U, V and W is achieved by simple matrix multiplication.

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = \begin{bmatrix} U_R & U_G & U_B \\ V_R & V_G & V_B \\ W_R & W_G & W_B \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Since the chromaticity co-ordinates define the proportion of each primary required to match a particular stimulus, the corresponding tristimulus values are simply the chromaticity co-ordinates multiplied by a constant.

$$\therefore \; T_{RGB \rightarrow UVW} = \begin{bmatrix} k_1 u_R & k_2 u_G & k_3 u_B \\ k_1 v_R & k_2 v_G & k_3 v_B \\ k_1 w_R & k_2 w_G & k_3 w_B \end{bmatrix}$$

where u_R, v_R and w_R are the chromaticity co-ordinates of the red primary, u_G, v_G and w_G are the chromaticity co-ordinates of the green primary and u_B, v_B and w_B those for the blue primary. Given the values of k_1, k_2 and k_3, it is possible to transform the U,V,W tristimulus values to R,G,B values using the inverse of the above matrix.

The values of k_1, k_2 and k_3 are derived by taking a particular case which, for television, by convention defines unit values of R,G,B to produce the chromaticity of a reference 'white'. For European television the reference white point is taken as the chromaticity matching illuminant D_{65} ($u = 0.1978$, $v = 0.3122$) and is produced when $R = G = B = 1$. Normalising the U,V,W tristimulus values so that the transformed luminance is also unity, the display of reference white D_{65} is described by the equation

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} T_{UVW \rightarrow RGB} \end{bmatrix} \begin{bmatrix} \dfrac{0.1978}{0.3122} \\ 1 \\ \dfrac{0.4900}{0.3122} \end{bmatrix}$$

Using a standard matrix inversion technique

$$T_{UVW \to RGB} = \begin{bmatrix} \frac{v_G w_B - v_B w_G}{k_1 \Delta} & \frac{u_B w_G - u_G w_B}{k_1 \Delta} & \frac{u_G v_B - u_B v_G}{k_1 \Delta} \\ \frac{v_B w_R - v_R w_B}{k_2 \Delta} & \frac{u_R w_B - u_B w_R}{k_2 \Delta} & \frac{u_B v_R - u_R v_B}{k_2 \Delta} \\ \frac{v_R w_G - v_G w_R}{k_3 \Delta} & \frac{u_G w_R - u_R w_G}{k_3 \Delta} & \frac{u_R v_G - u_G v_R}{k_3 \Delta} \end{bmatrix}$$

with $\Delta = u_R(v_G w_B - v_B w_G) + u_G(v_B w_R - v_R w_B) + u_B(v_R w_G - v_G w_R)$

For the RCA shadow mask CRT type A49-120X, the phosphors have chromaticities:

	u	v
Red	0.4169	0.3521
Green	0.1371	0.3741
Blue	0.1799	0.1097

Using these phosphor chromaticities and substituting the conditions for display of the reference white (D_{65}) gives the values:

$k_1 = 0.5876$

$k_2 = 1.9145$

$k_3 = 0.7010$

$$\text{hence} \quad T_{UVW \to RGB} = \begin{bmatrix} 4.9425 & -0.2201 & -1.2178 \\ -1.6069 & 1.8205 & 0.1259 \\ 1.6726 & -3.3611 & 2.1035 \end{bmatrix}$$

$$\text{and} \quad T_{RGB \to UVW} = \begin{bmatrix} 0.2450 & 0.2625 & 0.1261 \\ 0.2069 & 0.7162 & 0.0769 \\ 0.1357 & 0.9358 & 0.4980 \end{bmatrix}$$

In particular, $V = 0.2069R + 0.7162G + 0.0769B$.

3.52 Gamma correction.

Precise reproduction of chromaticity by additive mixture of light sources, as in television, requires that a linear relationship is maintained between the display light output and that in the original scene. Unfortunately, the television display tube is an inherently non-linear device. In general, the light output is related to the input grid voltage by a power law of the form

$L = AV^{\gamma}$ (where L is screen luminance, V grid voltage and A and γ are constants)

The transfer characteristic of a typical colour display tube is illustrated by Figure 3.13. This characteristic has an approximate γ of 2.2. Suppose that this tube were used, assuming its response to be linear, to reproduce the chromaticity $u = 0.22$, $v = 0.165$ with a luminance of 10 cd/m^2, the required red, green and blue channel luminances are calculated using the following procedure:

$$U = \frac{uV}{v} = \frac{0.22 \times 10}{0.165} = 13.33$$

$$V = 10$$

$$W = (1 - 0.22 - 0.165) \times \frac{U}{u} = 37.26$$

The red, green and blue luminances R,G,B corresponding to this chromaticity may be calculated from a transform matrix (see section 3.51):

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} \ T \ \end{bmatrix} \begin{bmatrix} U \\ V \\ W \end{bmatrix}$$

For a typical display, $[T]$ has values:

$$\begin{bmatrix} 4.943 & -0.220 & -1.218 \\ -1.607 & 1.821 & 0.126 \\ 1.673 & -3.361 & 2.103 \end{bmatrix}$$

giving R = 18.310 G = 1.475 B = 67.059

for a linear transfer characteristic. For a tube gamma of 2.2, the corresponding luminance signals become R = 599.67 , G = 2.35 , B = 10428.10.

Using the transform

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = \begin{bmatrix} T^{-1} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \qquad \text{(see section 3.51)}$$

gives values $U = 1462.44 \quad V = 927.39 \quad W = 5275.56$

$\therefore \quad u = 0.19 \qquad v = 0.12$

Thus, using a display tube gamma of 2.2, the required chromaticity of $u = 0.22$, $v = 0.165$ becomes $u = 0.19$, $v = 0.12$, a difference of 0.054. At the B.B.C. research department a chromaticity error of 0.00384 on the 1960 CIE-UCS diagram is considered significant under average viewing conditions, this being described as 1 just-noticeable difference (jnd). Using this criterion an error of 0.054 is unacceptable for accurate chromaticity reproduction. In general, increasing the display gamma always increases saturation in the direction of the dominant display primary. The luminance also shows a considerable increase, although this would be simply compensated for by reduction of chrominance gain. This increased saturation is exploited in broadcast television which uses an overall system gamma of $\sim$ 1.27 (B.B.C. Research Dept., 1971)[20].

It is known that a fall in luminance can produce a subjective loss of saturation (Hunt, 1965[18]; Bartleson and Witzel, 1967[19]). Since television displays have a relatively restricted luminance range, exact luminance reproduction is usually not possible. This lower reproduction luminance is compensated by a deliberate saturation increase resulting from incomplete gamma correction.

For a computer graphic display having exact chromaticity reproduction as a design requirement, any display tube non-linearities must be taken account of and compensated for. This compensation or gamma correction can be achieved using a device with a transfer characteristic which is the inverse of that of the display tube to predistort the video signal amplitude. Since the video signal is derived from the output data of a digital computer, it is convenient to use software gamma correction, allowing a change of display tube or display monitor to be quickly and accurately compensated for.

Figure 3.13 shows the transfer characteristics for a Philips colour monitor type EL8560, fitted with an RCA type A49-120X shadow mask tube. This characteristic is an almost exact power law with a γ of 2.1. Figure 3.14 shows the system transfer characteristic, from computer to display screen, using a software gamma correction of 0.476 along with characteristics for assumed γ of 1.9 and 2.2.

3.53 Colour discrimination and the television display.

The precision to which a particular display screen luminance level may be selected depends on the transfer characteristic, or "gamma" of the display tube as well as the type of quantisation used for the video signal in its digital form. Each colour channel is linearly quantised into 256 discrete levels, thus the discrimination of output chromaticity increases as the overall luminance is increased. Fortunately, the reduction of precision in chromaticity reproduction as luminance is reduced is accompanied by a reduction in the human visual system's ability to resolve small colour differences at low

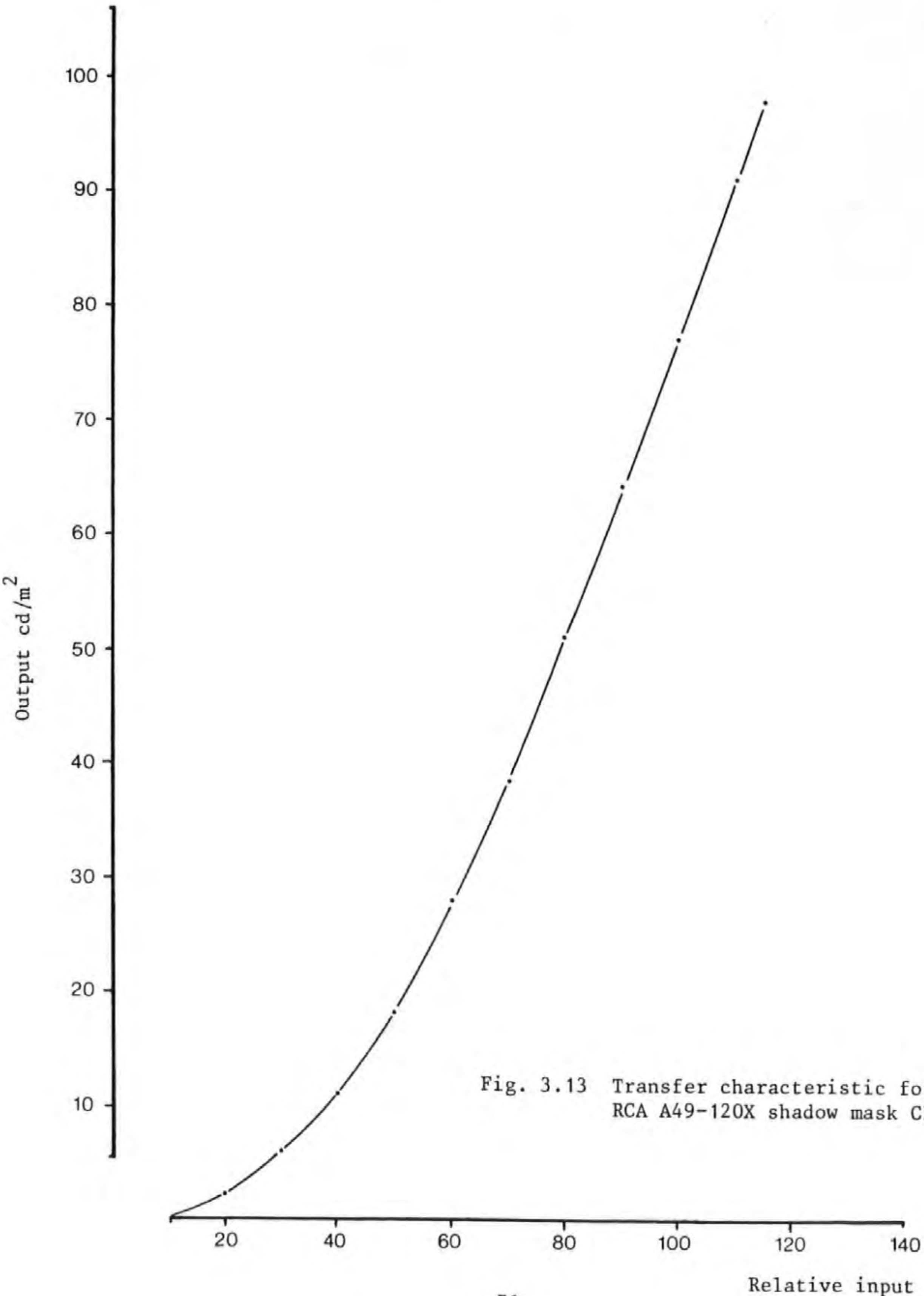


Fig. 3.13 Transfer characteristic for RCA A49-120X shadow mask CRT.

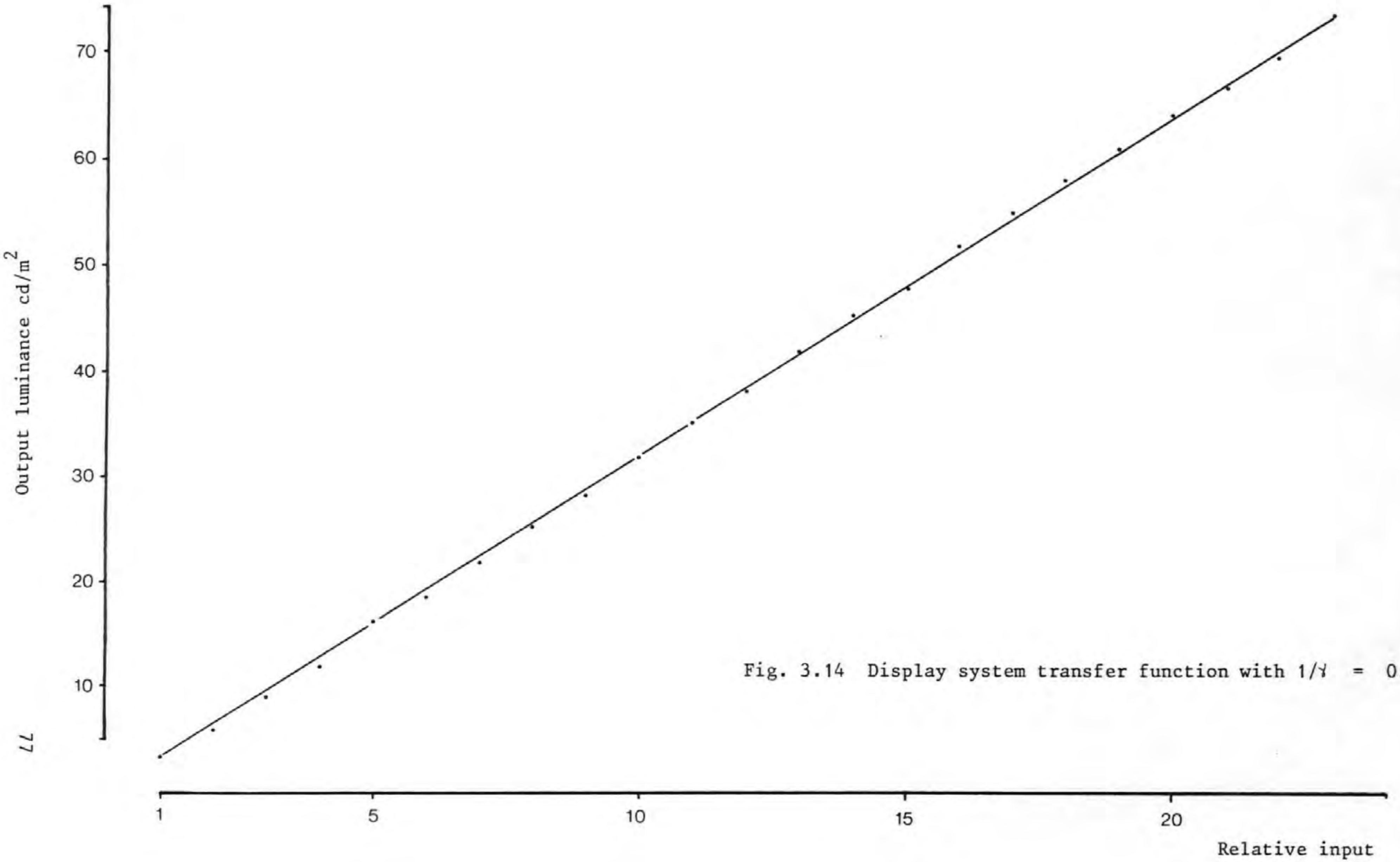


Fig. 3.14 Display system transfer function with $1/\gamma = 0.476$

luminance levels.

Figure 3.15 shows the distance in CIE uniform colour space equivalent to one unit in the CIE U',V',W' three dimensional colour space (see section 3.37) for a range of luminances. This CIE unit is chosen to be at or near the threshold of perceptibility (Judd and Wyszecki, 1963)[21]. Various studies have been conducted to find the limiting unit of colour discrimination or "just noticeable difference" (jnd) for different viewing conditions (MacAdam, 1942[22]; Wright, 1943[23]; Judd, 1936[24]). Work at the B.B.C. research department has shown, for television, an error of 0.00384 in the 1960 CIE UCS diagram to be significant for average viewing conditions, although errors as small as 0.0004 may be perceptible under good viewing conditions at high luminance (Wood and Sproson, 1971[17]).

Figure 3.16 shows the chromaticity resolution for the colour display corresponding to a one bit change in the red channel output whilst displaying a chromaticity close to D_{65}, over a typical range of output luminance. This resolution is dependent on the display tube which in this case has a gamma of 2.1. The correction for this CRT non-linearity reduces slightly the maximum colour discrimination. Comparison of Figures 3.15 and 3.16 shows that over the whole luminance range, the display chromaticity discrimination exceeds that of the viewer by a significant margin. The CIE unit of colour difference is based on the formula $W' = 25Y^{1/3} - 16$. For a television display, the brightness is related to luminance by a relationship $B \simeq aL^{\gamma}$ where a is constant and the power γ varies for $\sim 1/3$ for viewing with a dark surround to $\sim 1/2$ for viewing with bright ambient illumination. Thus, the CIE brightness

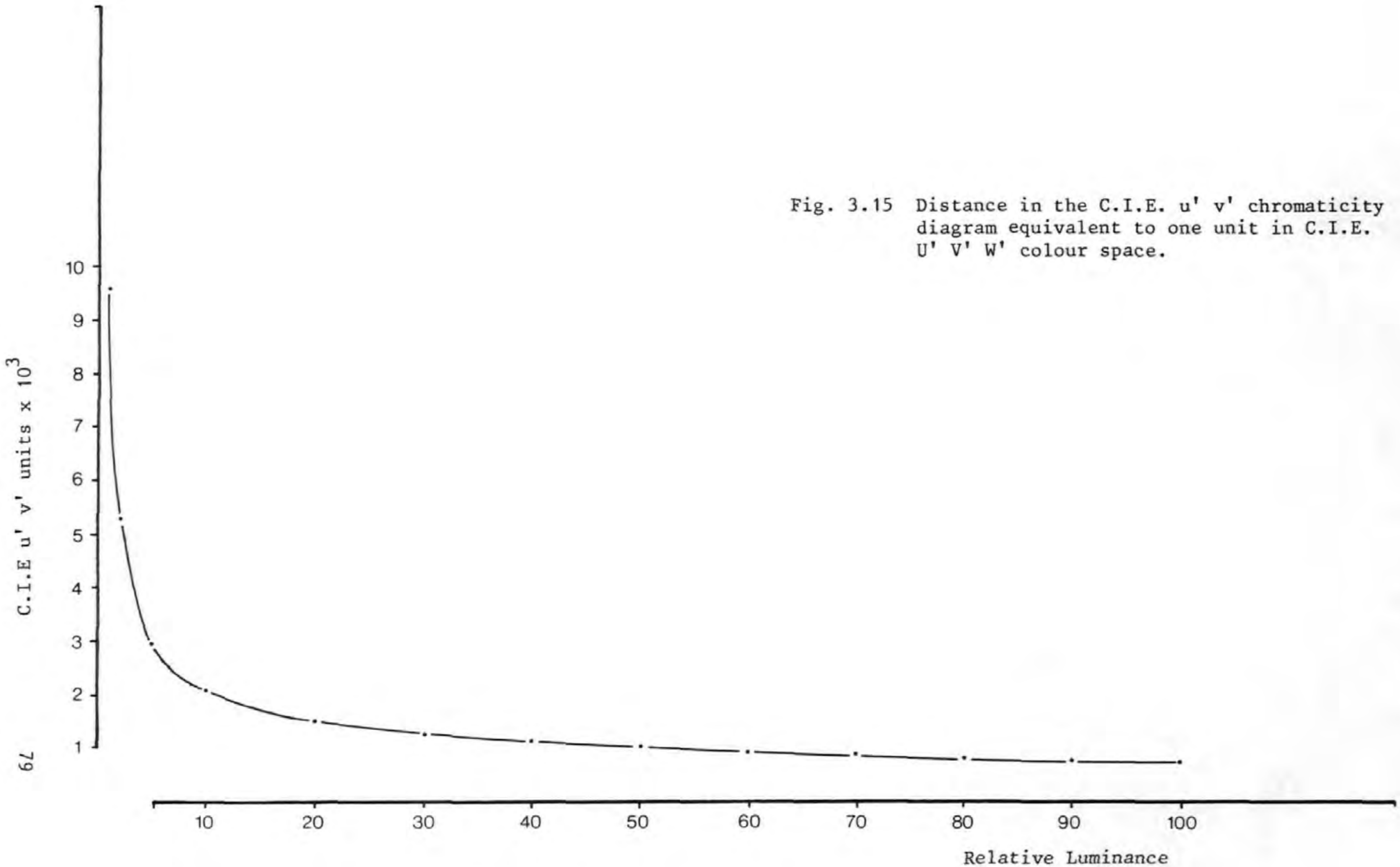


Fig. 3.15 Distance in the C.I.E. u' v' chromaticity diagram equivalent to one unit in C.I.E. U' V' W' colour space.

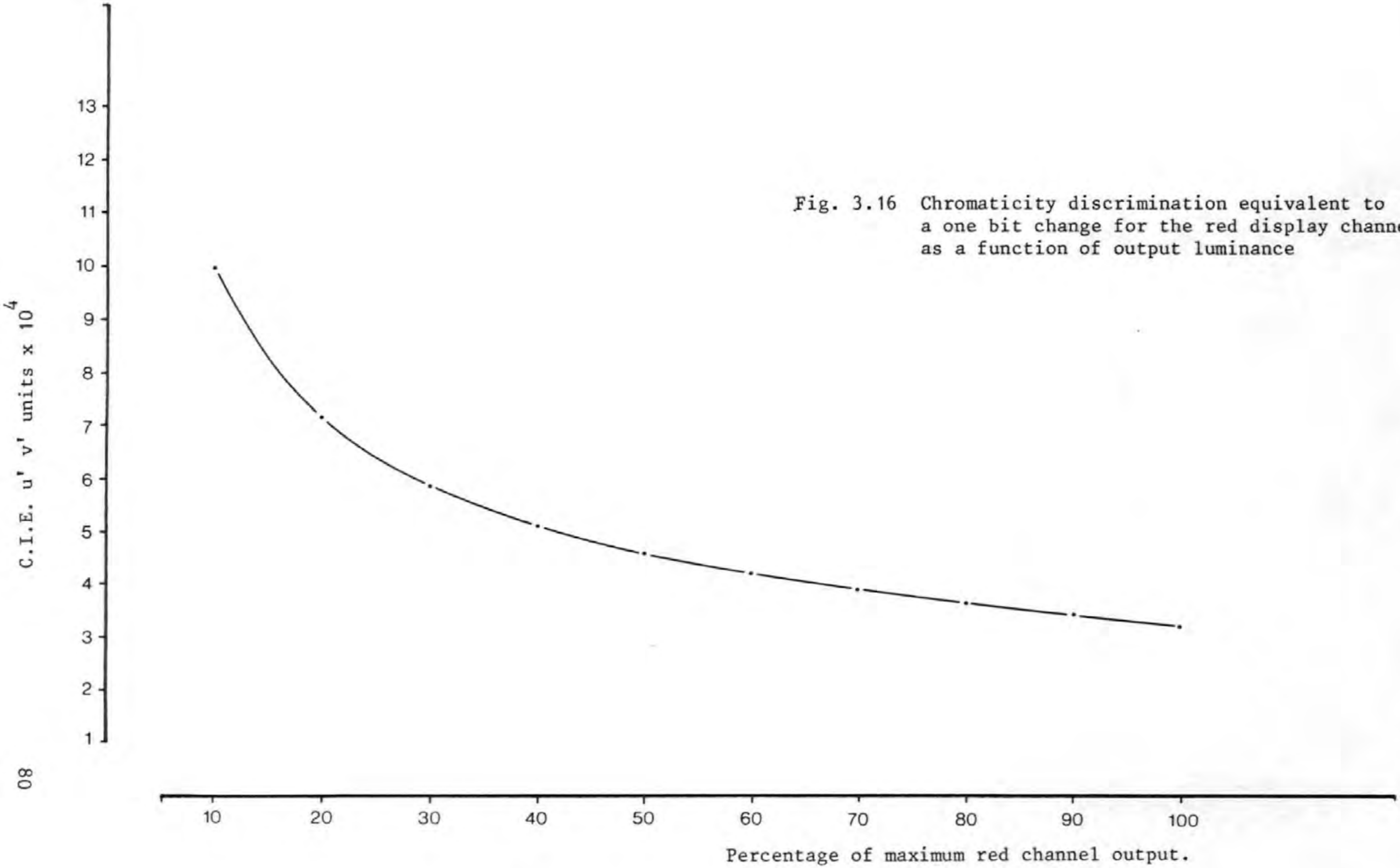


Fig. 3.16 Chromaticity discrimination equivalent to a one bit change for the red display channel as a function of output luminance

unit is most appropriate when viewing television with a dark surround.

3.54 Photometric measurement techniques.

3.54.1 Introduction.

Accurate reproduction of chromaticity using a television monitor requires detailed knowledge of the characteristics of the display tube. The discussion of 3.41 shows that the chromaticities of the three colour phosphors must first be found. Secondly, the grey scale tracking of the monitor must be adjusted such that the required reference 'neutral' point (in this case D_{65}, u = 0.1978, v = 0.3122) is reproduced whenever the three display inputs R,G,B are set equal, for all levels of input signal. As discussed in section 3.42, the inherent non-linearity of the display tube must be quantified so that a 'gamma' correction may be applied. The following sections describe the photometric techniques used to measure these quantities and to provide a means of periodically checking the system accuracy. A guide to the system setting-up procedure is given in Appendix 3A. For a review of the effects of errors in the setting-up of a display system, the reader is referred to Pearson (1975)[25].

3.54.2 Measurement of the display-screen luminance.

Routine measurement of screen luminance is carried out using a Hagner photometer model S2. The instrument uses silicon diode light sensors preceded by optical filters so that the overall spectral response of the instrument is closely related to that of the CIE standard observer for a two degree field. The high attenuation of the filters is compensated for by a linear high gain amplifier. As shown in Figure 3.17, the spectral response function of the photometer does not provide a perfect match for that of the standard observer. Thus the instrument will produce corresponding errors, the magnitude of which depend on the spectral power distribution of the stimulus being measured. Spectral power distributions of the display tube primary phosphors are shown in Figures 3.18 - 3.20. In order to measure the display screen luminance, the photometer accuracy was checked using a calibrated lamp and a photometer bench. The experimental arrangement is shown in Figure 3.21.

The standard lamp was first adjusted to give a colour temperature of 2700°K. The television monitor was then adjusted to match this colour temperature using the photometer head. To achieve a balance of the intensities of the two light sources, calibrated neutral density filters were used to attenuate the output from the lamp. The screen was masked, leaving a central area of one square inch as the light source, providing a convenient measure of screen luminance. Finally, the photometer head was positioned for photometric balance, allowing the intensity of the display screen aperture in candelas and hence the screen luminance in candelas/sq.inch to be calculated.

Having measured the screen luminance, it remains to calibrate the Hagner photometer. Three correction factors are derived, one for each of the television primary colours. From a knowledge of the three primary chromaticities and the chromaticity of the reference neutral, the amounts of each primary required to produce, say, 1 apostilb of 'white' can be found as described in section 3.51. From this information

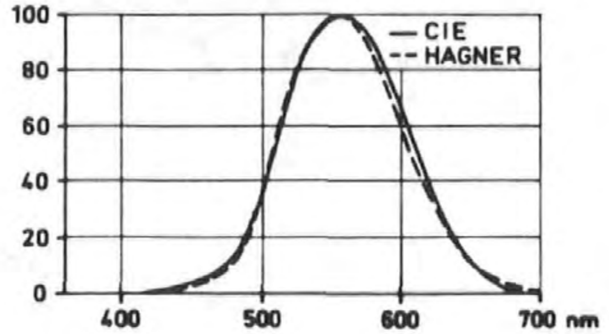


Fig. 3.17 Hagner photometer response compared with the C.I.E. luminosity curve.

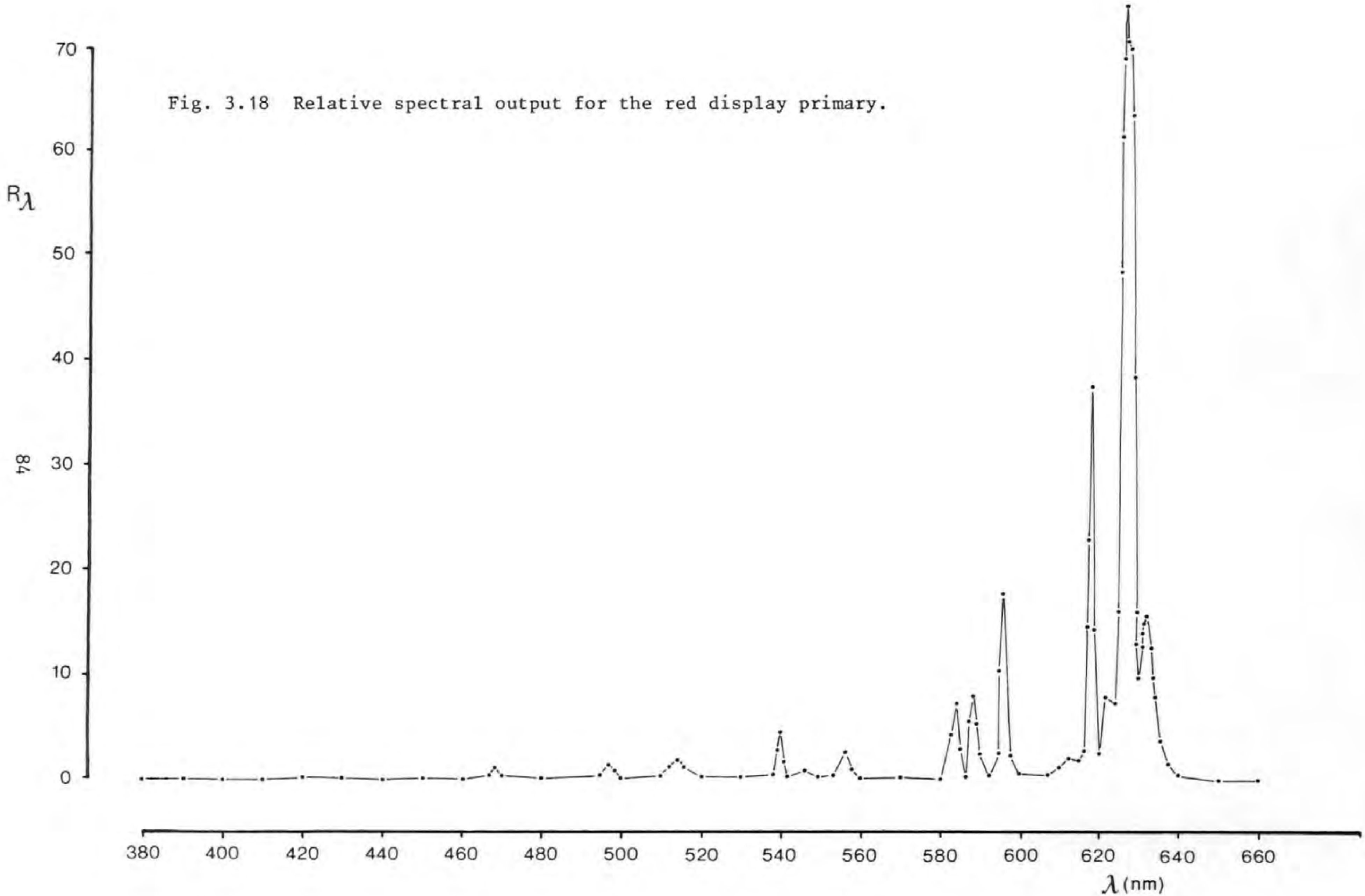
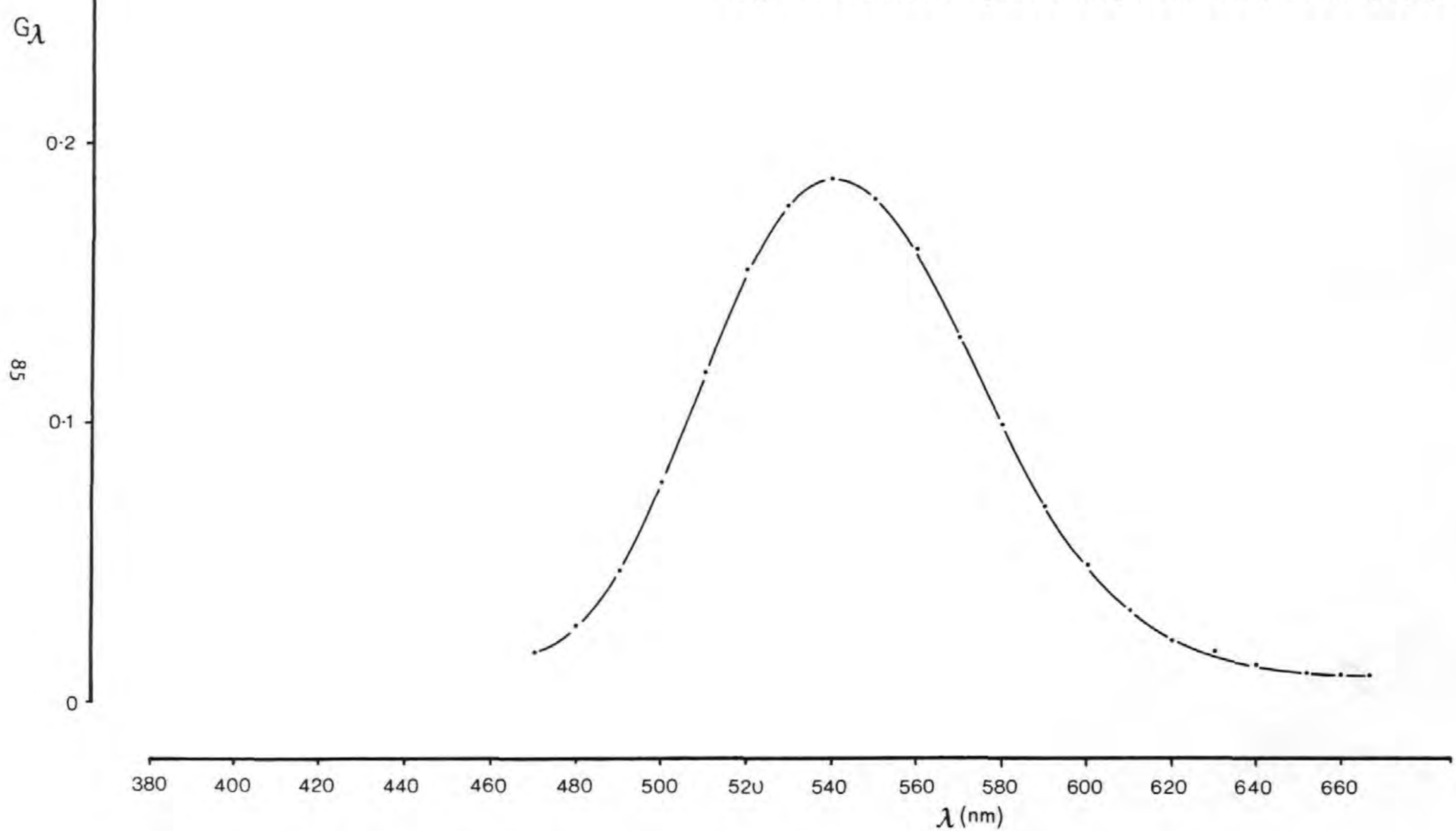


Fig. 3.18 Relative spectral output for the red display primary.

Fig. 3.19 Relative spectral output for the green display primary.



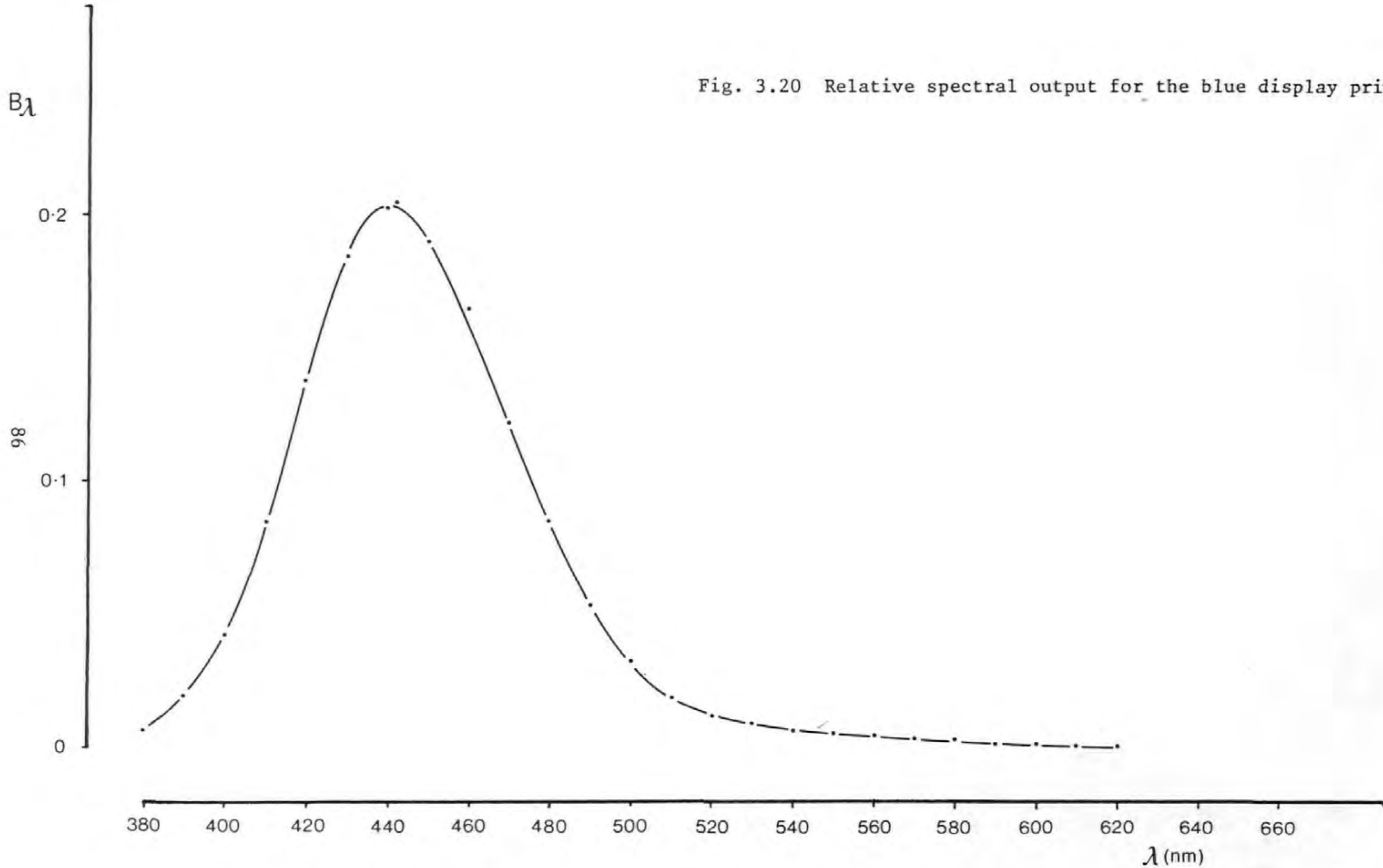


Fig. 3.20 Relative spectral output for the blue display primary.

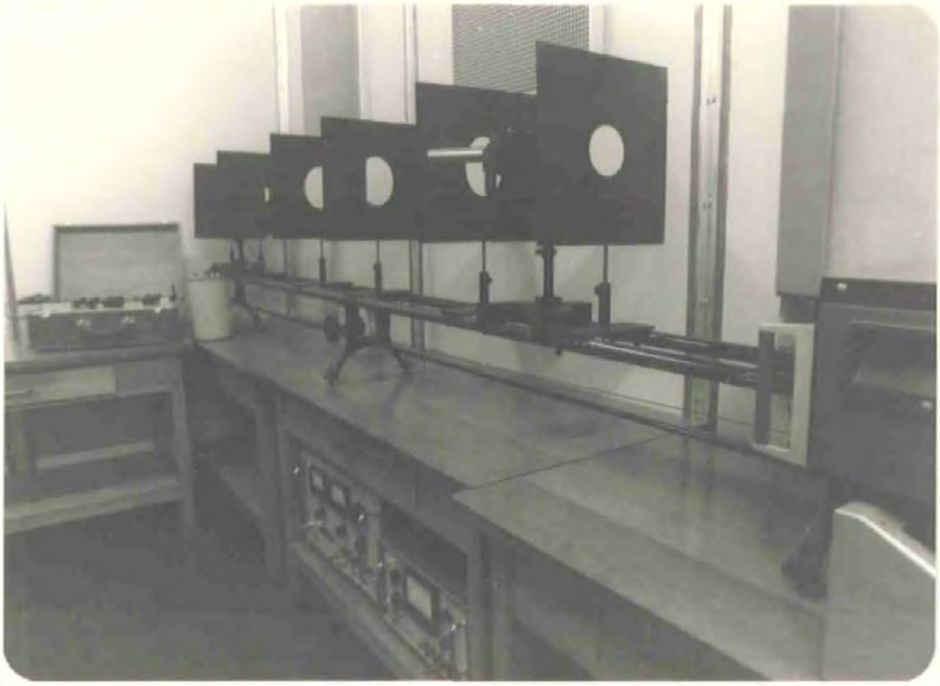


Fig 3.21 Experimental arrangement for photometer calibration.

the output of each primary phosphor required to match 1 apostilb of some other colour, for example, that corresponding to a colour temperature of 2700°K, may be found. In the case of the Philips monitor used for this work the relevant chromaticities are:-

	x	y
RED PRIMARY	0.620	0.349
GREEN PRIMARY	0.320	0.584
BLUE PRIMARY	0.155	0.063
REFERENCE NEUTRAL (D_{65})	0.3127	0.3291

For these chromaticities the total luminance Y is given:-

$$Y = Y_R + Y_G + Y_B$$

$$= 0.2068R + 0.7164G + 0.0768B$$

for a luminance of 1 apostilb at a colour temperature of 2700°K (a chromaticity of x = 0.4595 y = 0.4105).

$$\left.\begin{array}{l} Y_R = 0.4406 \\ Y_G = 0.5442 \\ YB = 0.0152 \end{array}\right\} \text{apostilbs}$$

for an accurately measured display screen luminance Y

$$\therefore \quad Y_R = 0.4406Y \quad Y_G = 0.5442Y \quad Y_B = 0.0151Y$$

If the readings for the three primaries, measured separately by the Hagner photometer are respectively R_H, G_H and B_H

$$\therefore \quad R_H \times k_R = 0.4406Y$$

$$G_H \times k_G = 0.5442Y$$

$$B_H \times k_B = 0.0151Y$$

hence $k_R = \dfrac{0.4406Y}{R_H}$

$$k_G = \frac{0.5442Y}{G_H} \quad \text{etc.}$$

Thus by measuring the screen luminance on a photometer bench, and the respective luminance readings of the three primary components using the Hagner photometer, three correction factors are obtained.

For a screen luminance of 57.67 cd/m² at a colour temperature of 2700°K

$$\left.\begin{aligned} R_H &= 22 \\ G_H &= 29 \\ B_H &= 0.6 \end{aligned}\right\} \text{cd/m}^2$$

$$\therefore \quad k_R = \frac{0.4406 \times 57.67}{22} = 1.1549$$

similarly $k_G = 1.0823$

$$k_B = 1.46$$

The corrected screen luminance is obtained by multiplying readings from the Hagner photometer for each phosphor separately by the appropriate constant.

$$\therefore \quad L = k_R R_H + k_G G_H + k_B B_H$$

3.54.3 Measurement of chromaticity using a monochromator.

Two methods have been recommended by the International Electrotechnical Commission (I.E.C., 1974)[26] for the measurement of the chromaticity of cathode-ray tube phosphors. One approach uses a calibrated spectrometer or monochromator which measures the spectral energy distribution of the light output, wavelength by wavelength. From this information the chromaticity may be readily calculated. The technique can be more accurate than other methods but involves a rather lengthy series of measurements. In the early stages of this work, the method was adopted generally for chromaticity measurement, largely because of its potentially high accuracy and the availability of a suitable monochromator.

The monochromator (a Spex 'Minimate') has as its central component a precision diffraction grating, mounted on a rotatable shaft which is mechanically coupled to a wavelength-calibrated dial. The wavelength may be adjusted manually or by a small synchronous motor which allows the use of a graph plotter. Incoming light passes through a slit and is then focussed by mirrors on to the diffraction grating. The analysed light is focussed on to a further slit from which it passes to a suitable light sensor. In this case the sensor was a photomultiplier tube type IP28A/V1.

3.54.4 Calibration of the monochromator.

The wavelength scale calibration is checked by comparing the wavelength reading from known spectral emission lines. A convenient source of known frequencies is provided by a fluorescent tube, the output of which has a number of monochromatic lines associated with mercury vapour. The bandwidth may be measured by adjusting the monochromator to points on either side of a known line peak, at which half peak output is obtained, noting the difference in wavelength between the two points. The bandwidth is determined largely by the width of the input and output slits, but varies across the spectrum, widening as the wavelength increases. This variation, combined with the spectral response of the photomultiplier (see Figure 3.22) and the absorption characteristics of the optical paths in the monochromator, leads to a variation of sensitivity with wavelength. By measuring the response to a source of illumination whose spectral characteristics are known, this variation may be expressed numerically.

For the Spex Minimate and an RCA IP28A/V1 photomultiplier, the correction function $F(\lambda)$ was derived using a sub-standard tungsten-filament light source. The tungsten filament lamp provides a spectral output closely resembling that for a black body radiator, for which spectral distributions for given colour temperatures are well documented (Wyszecki and Stiles, 1967)[27]. The sub-standard lamp used for this work was calibrated to give an output at 2700°K, corresponding to a current of 3.467 amps. If $S(\lambda)$ is the spectral power distribution of a black body of 2700°K and $R(\lambda)$ is the response produced by the monochromator system, then $F(\lambda) = R(\lambda)/S(\lambda)$. For the above system the function $F(\lambda)$ is plotted in Figure 3.23. Knowing $F(\lambda)$, the system may be used to determine the spectral power distribution of other sources of visible radiation.

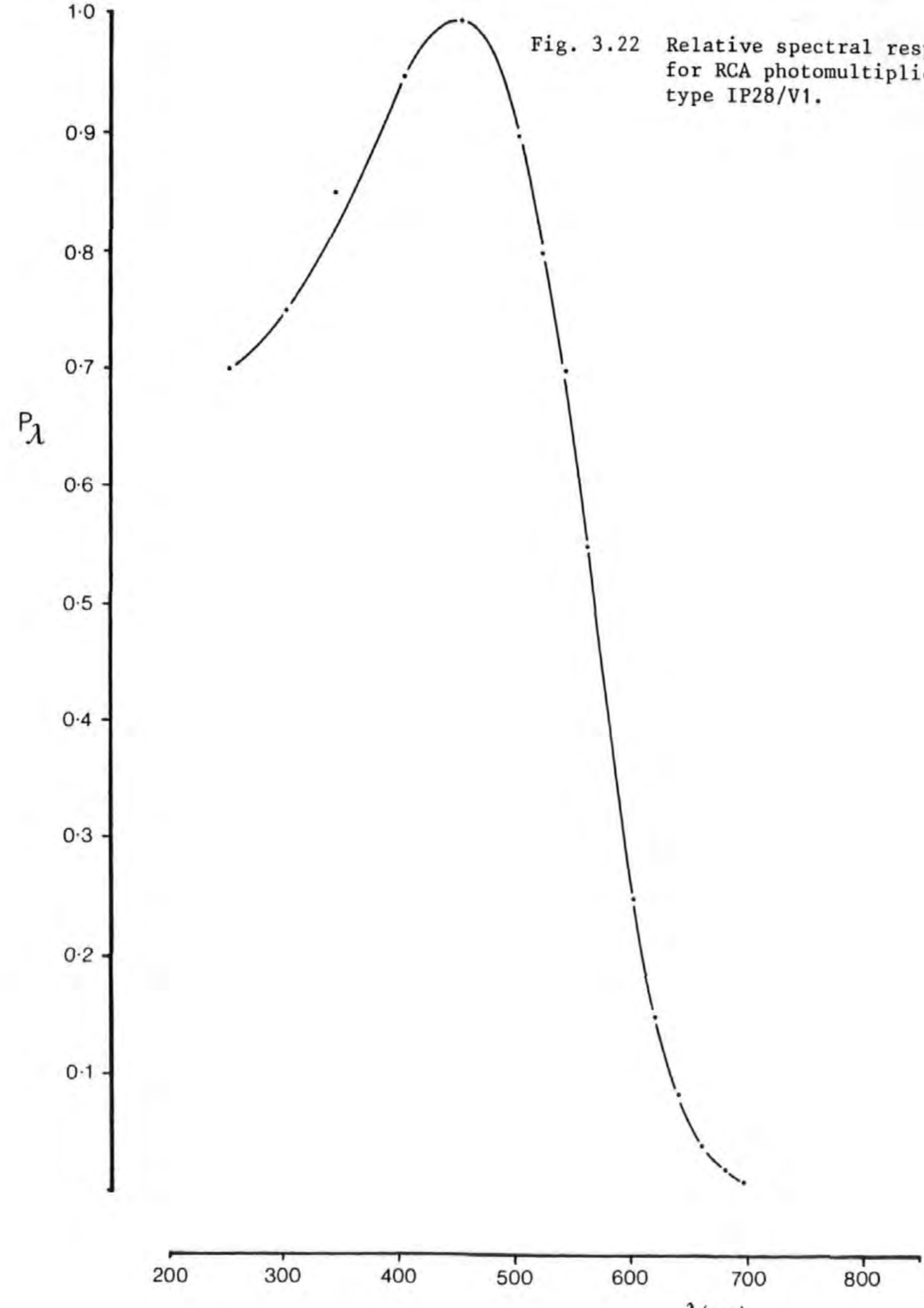


Fig. 3.22 Relative spectral response for RCA photomultiplier type IP28/V1.

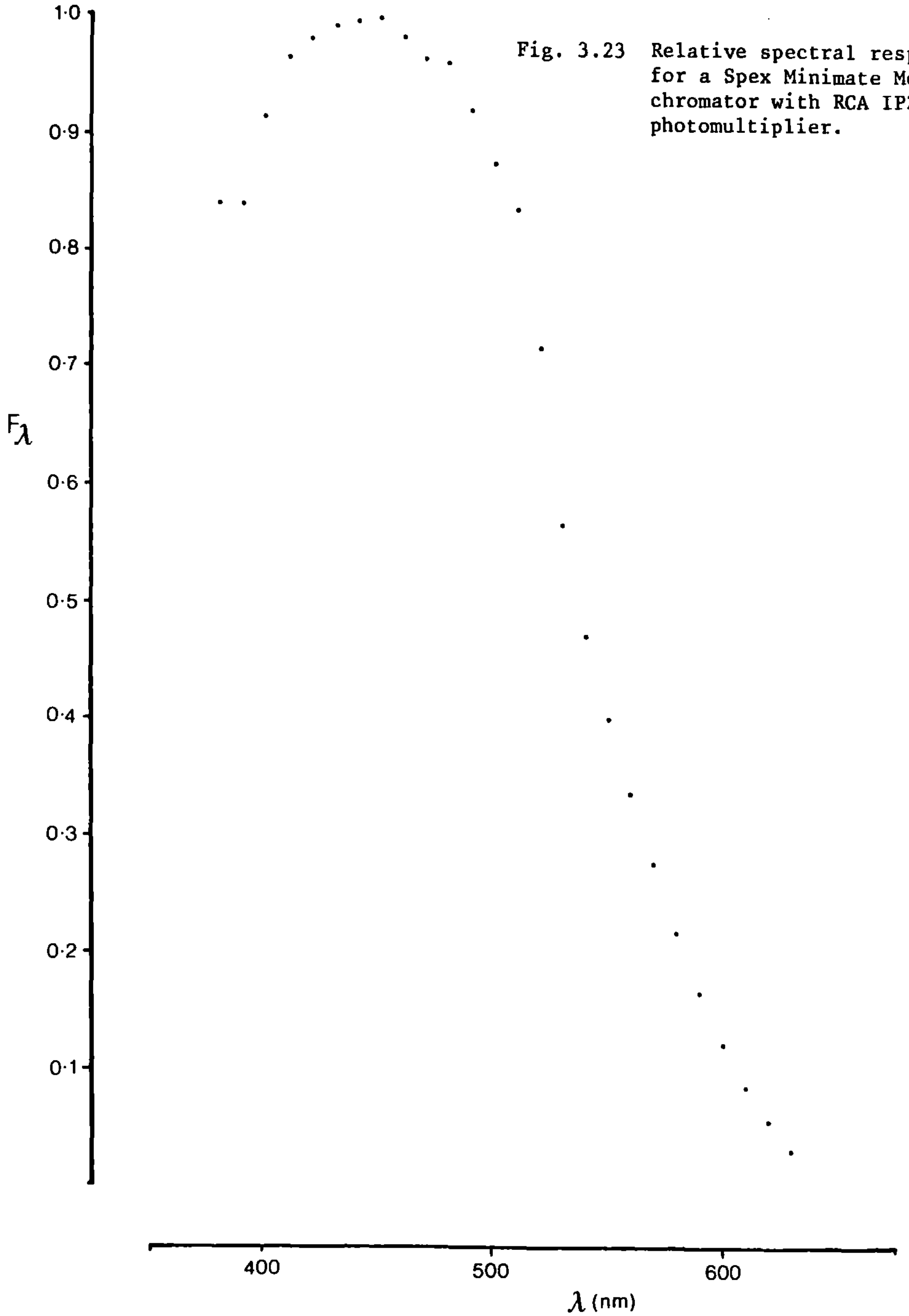


Fig. 3.23 Relative spectral response for a Spex Minimate Monochromator with RCA IP28/V1 photomultiplier.

3.54.5 Calculation of chromaticity co-ordinates.

Using the 1931 C.I.E. colour space, the tristimulus values X,Y,Z for a particular stimulus are defined by equations of the form:

$$X = \sum_{\lambda=380\ nm}^{\lambda=780\ nm} R(\lambda)/F(\lambda)\ .\ \bar{x}(\lambda)\Delta\lambda$$

with similar equations for Y and Z where $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ respectively are substituted for $\bar{x}(\lambda)$ where

$R(\lambda)$ = output reading at wavelength λ

$F(\lambda)$ = correction function at wavelength λ

$\bar{x}(\lambda),\bar{y}(\lambda),\bar{z}(\lambda)$ = CIE tristimulus spectral distribution coefficients for an equal energy stimulus at wavelength λ

$\Delta(\lambda)$ = equal wavelength interval between measurements

The corresponding x, y and z chromaticity co-ordinates may be calculated from X, Y and Z since

$$x = \frac{X}{X + Y + Z} \qquad y = \frac{Y}{X + Y + Z} \qquad z = \frac{Z}{X + Y + Z}$$

Choice of the wavelength interval $\Delta\lambda$ is governed by the nature of the spectral distribution being measured. Figures 3.18–3.20 show typical spectral power distributions for the outputs from the red, green and blue phosphors used in colour television cathode-ray tubes. The closely spaced high-energy narrow-band components produced in the far-red regions of the red phosphor spectrum can only be resolved by closely spaced measurements using an instrument with a narrow passband. Conversely, the blue and green phosphors give a mainly continuous distribution of energy, making detailed narrow-band measurements unnecessary. For blue and green phosphors the measurements can thus be made using larger wavelength intervals, the shorter period of time required for measurement reducing the possibility of drift errors. A further discussion of the sources of error inherent in this technique is given in Appendix 3B.

3.54.6 Colour measurement using a colorimeter.

The disadvantages of colour measurement using a monochromator were outlined above and discussed further in Appendix 3B. Some of these problems are avoided by use of a colorimeter. Such a device* is shown in use in Figure 3.24 (Philippart, 1966)[28]. The colorimeter offers the major advantage that chromaticity may be measured within a relatively short time, only six readings being required to quantify each colour. Currently, more sophisticated devices are available which allow direct digital readout of chromaticity (Hacking, 1976)[29].

The colorimeter measures the three 'tristimuli' X,Y,Z as specified for the C.I.E. standard observer, the stimulus being sampled over a two degree field. This allows measurement of small areas of a television

*Kindly provided by the B.B.C. Research Dept., Kingswood Warren.

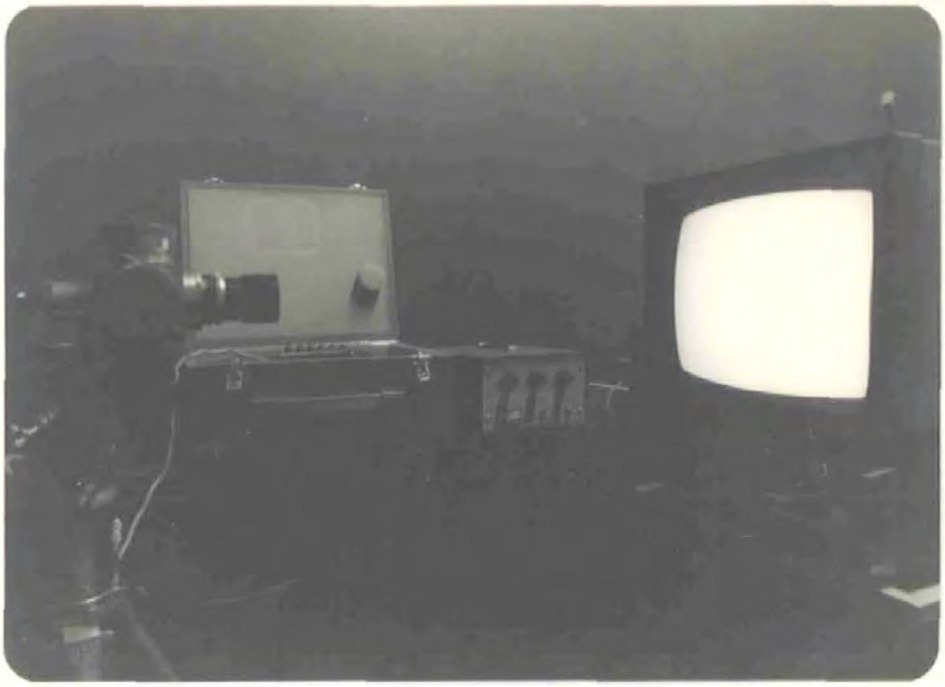


Fig. 3.24 Colour measurement using a colorimeter.

picture at a convenient viewing distance. A single photomultiplier tube is used as the light receptor, in conjunction with a set of colour filters. The filters are chosen to provide, in conjunction with the spectral response characteristics of the photomultiplier tube and the absorption characteristics of the lenses and mirrors, a close match with the spectral responses $\bar{x}$, $\bar{y}$ and $\bar{z}$ of the C.I.E.. 2° standard observer. The closeness with which the instrument matches these functions is shown in Figure 3.25. As can be seen, the $\bar{x}$ and $\bar{y}$ responses are too high at longer wavelengths due to the response characteristic of the photomultiplier tube. To overcome this difficulty, two additional filters are used, which are adjusted in response and density to give readings, which when subtracted from the X and Y values respectively, permit correct tristimulus values to be obtained.

To derive the values X,Y,Z the instrument uses six filters, mounted so that they may be successively placed in front of the photomultiplier tube. The filters give readings of X_R (the 'red' passband of X), X_B (the 'blue' passband of X), Y, Z, -X' and -Y', the latter two filters provide correction for the X and Y readings as described above. Further details of the design of the instrument are given by Philippart (1966)[28].

3.54.7 Colorimeter calibration.

Due to variations between individual photomultiplier tubes of the same type, some form of magnitude adjustment is required for the output of each filter. This correction is made electronically using potentiometers since 'neutral' density filters

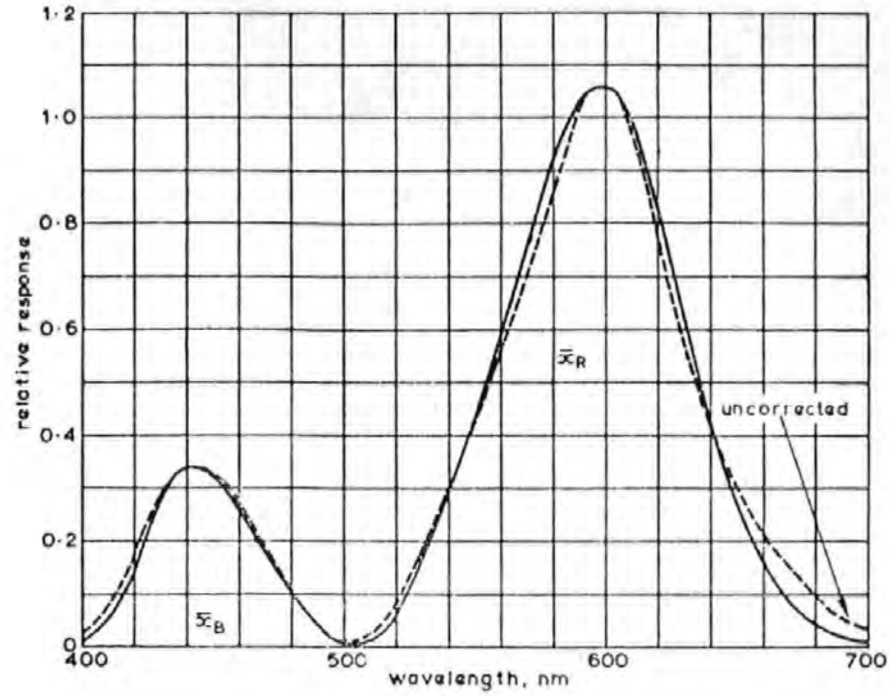


Fig. 3.25(a) Colorimeter filter approximation to the C.I.E. x colour matching function.

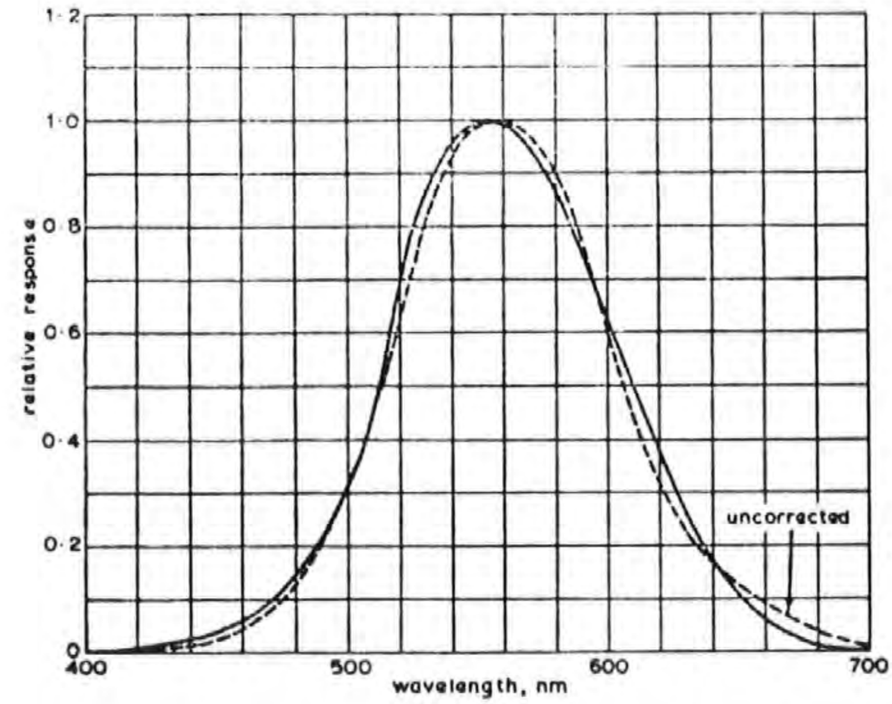


Fig. 3.25(b) Colorimeter filter approximation to the C.I.E. y colour matching function.

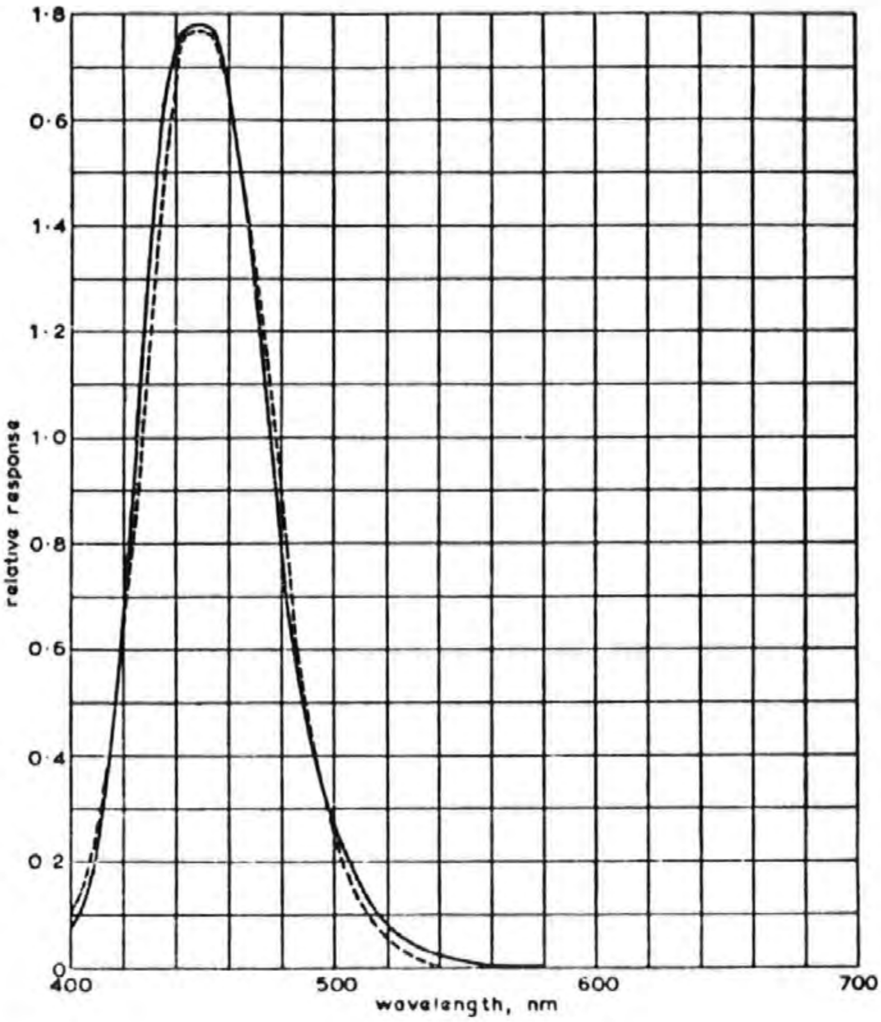


Fig. 3.25(c) Colorimeter filter approximation to the C.I.E. z colour matching function.

seldom have an ideal spectral characteristic. The potentiometers provide an adjustment range of ± 5%. Before use, the instrument was calibrated using a standard tungsten lamp of known colour temperature and a series of narrow-band colour filters. These filters were calibrated using a spectrophotometer, allowing the chromaticity co-ordinates of their response to the standard tungsten lamp to be accurately calculated.

A white tile having a neutral reflectance characteristic was illuminated by the standard lamp at a colour temperature of 2700°K. The colorimeter head was focussed on the tile and readings taken for the six internal filters respectively. By adjustment of the appropriate potentiometers and repeated readings, correct values of chromaticity could be obtained. In order to check calibration throughout the colour gamut, several colour filters were used. These are listed in Figure 3.26, together with typical readings produced by the colorimeter. The measurements are consistent with the claim by Philippart of a maximum error within 0.008 units in the CIE UCS chromaticity diagram.

Using this instrument, readings of tristimulus values can be obtained within one minute, compared with 20-30 minutes necessary when using a monochromator. Thus, errors introduced by drift within the display monitor are greatly reduced and a large number of colour stimuli may be quantified in a short time.

Light source	Chromaticity		Colorimeter readings	
	x	y	x	y
2700°K Standard lamp No filter	0.460	0.411	0.461	0.412
Standard lamp + 'Wratten' W25 Red filter	0.677	0.323	0.677	0.321
Standard lamp + 'Wratten' W58 Green filter	0.252	0.700	0.257	0.716
Standard lamp + 'Wratten' W47 Blue filter	0.141	0.056	0.139	0.050
Standard lamp + 'Wratten' W32 Magenta filter	0.572	0.240	0.569	0.241
Standard lamp + Ilford I303 Cyan filter	0.103	0.340	0.095	0.333
Standard lamp + CTR filter	0.3078	0.3237	0.306	0.316

Figure 3.26 a

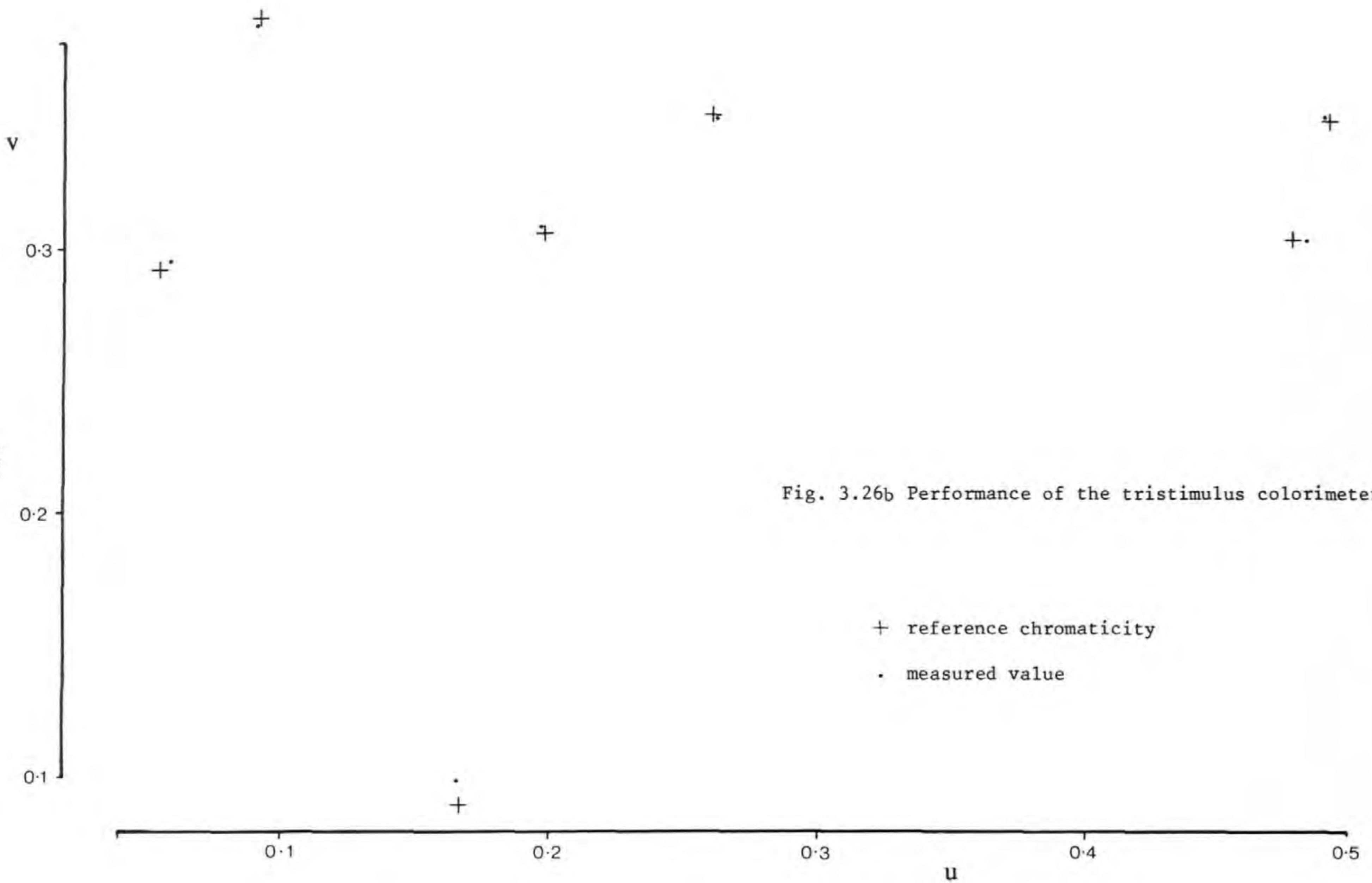


Fig. 3.26b Performance of the tristimulus colorimeter.

3.55 Accuracy of chromaticity reproduction.

The techniques described in section 3.54 provide an accurate measure of the display phosphor chromaticities, allowing the method of section 3.51 to be used to calculate the luminances L_R, L_G and L_B required to match any chromaticity within the display gamut. Before output to the display, these quantities are gamma corrected (see section 3.52) to preserve the linearity of the system. If the setting-up preliminaries have been carried out as outlined in Appendix 3A, the displayed chromaticity will closely match the input data.

The accuracy of chromaticity reproduction was checked using the colorimeter described in section 3.54, all chromaticity measurements being carried out in a photographic darkroom. The instrument was positioned at a distance of approximately two feet from the display screen giving, for a 2^o acceptance angle, a measured screen area of approximately 25 mm x 25 mm. Figure 3.27 shows input chromaticities (dots) plotted with their corresponding measured output chromaticities (crosses) in the 1960 CIE UCS diagram. These measurements are taken with a display screen luminance of 40 cd/m^2.

3.56 Chromaticity errors produced by ambient illumination.

The chromaticity reproduction performance of a display system is usually measured with the display in total darkness. Results from such measurement for the display system are shown in Figure 3.27a. In practice, television displays are often operated with substantial levels of ambient illumination. This ambient light is scattered from the display screen, usually causing a desaturation of the displayed colours.

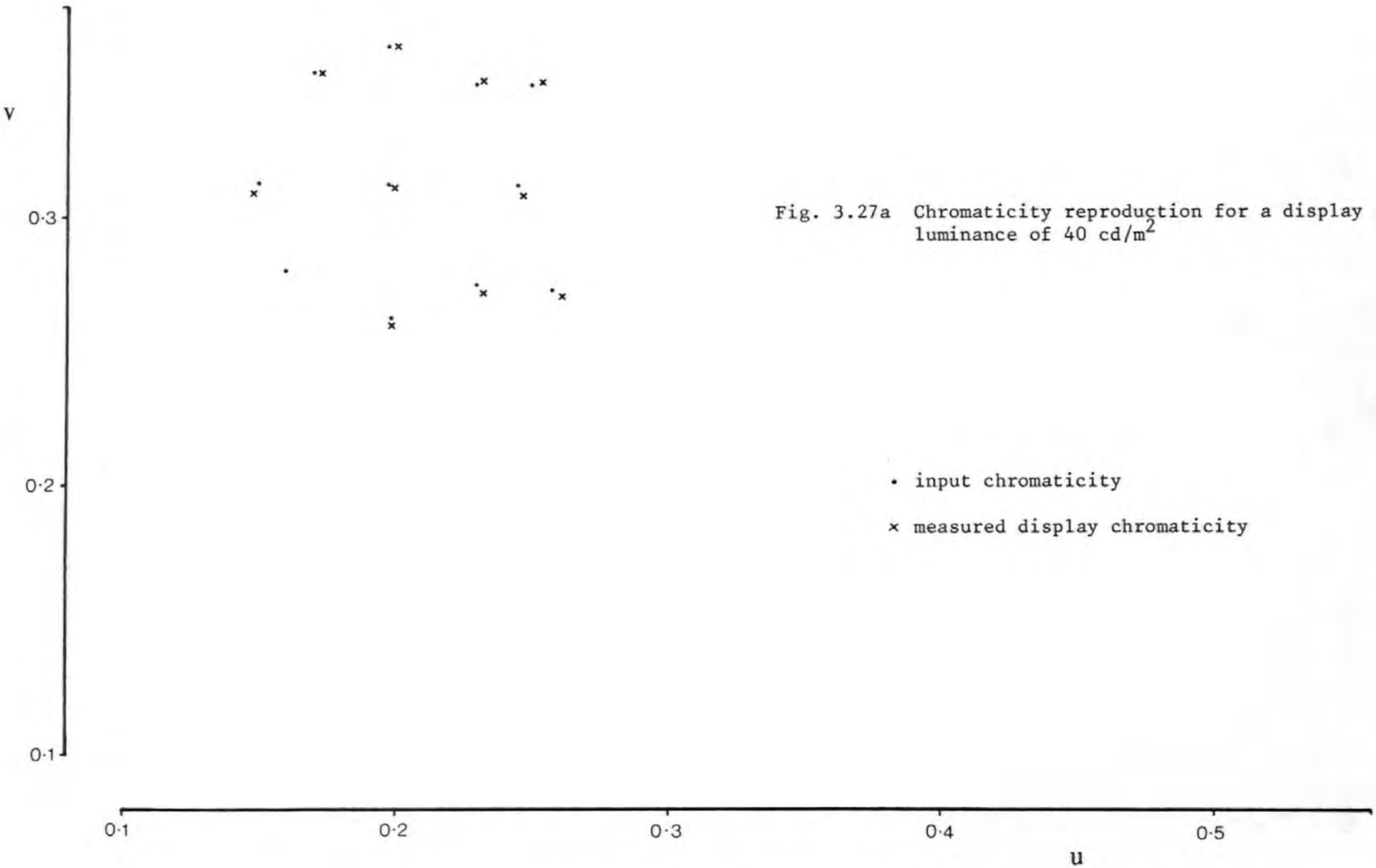


Fig. 3.27a Chromaticity reproduction for a display luminance of 40 cd/m^2

The effect of an ambient illuminance of 400 lux in the plane of the display screen is shown in Figure 3.27b for a 'Warm White' fluorescent lamp. The chromaticity shift is in the direction of the lamp chromaticity and, as would be expected, increases as the display luminance is decreased. The saturation loss may be partially offset by an increase in chrominance gain.

3.6 Colour applied to luminance design.

3.61 The general interreflection case.

In chapter 2, an interreflection model was described which, for a rectangular room, allows the direct and indirect components of illuminance, for each surface, to be quantified separately. It has been suggested by Lynes (1957)[1] and elsewhere (Joblove and Greenberg, 1978)[14], that interreflection between coloured surfaces may also be simulated using an extension of this technique.

The reflection characteristics of a Lambertian surface are defined by the spectral distribution of its reflectance. As discussed in section 3.54.5, it is normal practice in colorimetry to express spectral distributions as a series of narrow bands $\Delta\lambda$ within which the variable is assumed constant. In general, the use of 40 bands, each 10 nm wide, will provide sufficient accuracy. The colour characteristics of the light source may be defined using a spectral power distribution, also using a 40 band approximation.

Having defined the geometrical distribution of illumination using, for example, flux fraction ratio and BZ number, the luminance of

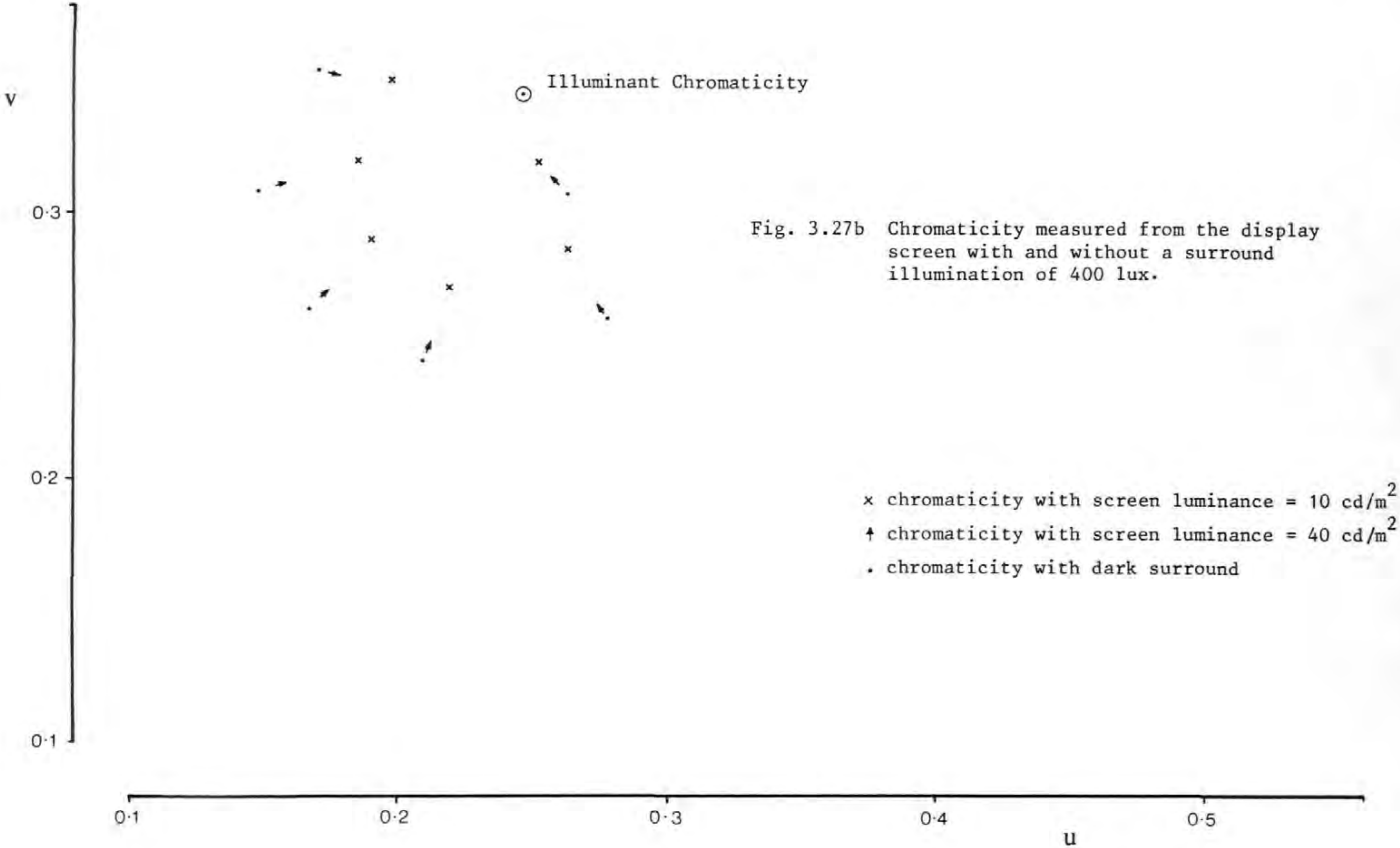


Fig. 3.27b Chromaticity measured from the display screen with and without a surround illumination of 400 lux.

each surface for each 10 nm band is found by solution of six simultaneous equations as described in chapter 2. Repeating this calculation forty times gives the spectral distribution of luminance for each surface and hence allows the resultant chromaticity and luminance for each surface to be obtained.

Using the above method, surface chromaticity is obtained which takes into account the chromatic interreflection from other surfaces as well as the colour rendering properties of the light source. This approach has two disadvantages. Firstly, the calculation time, of the order of one minute for the PDP11-10 computer using BASIC, is rather long for interactive use, although for many purposes this delay may be quite acceptable. Secondly, each surface colour and each light source requires storage of forty numbers. This may prove an unacceptable restriction for some computers, particularly where disc storage is not available.

3.62 An alternative approach to colour interreflection.

Although introducing some restrictions, a more direct approach to the colour interreflection calculation is available, offering reductions both in calculation time and required storage space. The restriction concerns the way in which the colour properties of the light source and the surfaces are defined.

The chromaticities of 88 surface colours, directly illuminated by D_{65}, are documented in British Standard BS 4800 (Supplement 1, 1975) which is described further in section 3.63. If the light source is confined to D_{65}, colour rendering is implicitly taken account of, leaving only the effect of interreflection to be calculated.

Each surface colour is assumed to be illuminated by three lights having respectively the chromaticities of the red, green and blue phosphors of the television display. The possibility of using a three-primary approximation has been suggested elsewhere (Joblove and Greenberg, 1978)[14]. Three 'neutral' reflectance coefficients ρ_R, ρ_G and ρ_B are assigned to each surface colour. The surfaces are assumed to reflect the red, green and blue lights without modifying their spectral characteristics. Thus, surface chromaticity, in the context of the television display, is defined by ρ_R, ρ_G and ρ_B.

Given the chromaticity u_s, v_s and reflectance ρ_s of a surface colour, the values ρ_R, ρ_G and ρ_B are obtained by the following procedure.

Suppose that the luminances L_R, L_G and L_B of the respective television primaries add to give the equivalent of 1 unit of surface luminance with a chromaticity matching that of D_{65}. From the discussion of section 3.3 these luminance values are respectively

$$L_R = 0.2069$$

$$L_G = 0.7162$$

$$L_B = 0.0769$$

if l_R, l_G and l_B are the three luminances of the simulated surface then

$$\frac{l_R + l_G + l_B}{L_R + L_G + L_B} = \rho_s = V_s$$

and

$$\begin{bmatrix} l_R \\ l_G \\ l_B \end{bmatrix} = \begin{bmatrix} \\ T_{UVW \to RGB} \\ \\ \end{bmatrix} \begin{bmatrix} U_s \\ V_s \\ W_s \end{bmatrix}$$

when U_s, V_s and W_s are the tristimulus values of the light reflected from the surface and $T_{UVW \to RGB}$ is the transform matrix derived in section 3.3. The values of ρ_R, ρ_G and ρ_B are now easily obtained since

$$\rho_R = {}^{l_R}/_{L_R} \qquad \rho_G = {}^{l_G}/_{L_G} \qquad \rho_B = {}^{l_B}/_{L_B}$$

Values of ρ_R, ρ_G and ρ_B are derived for each interreflecting surface. The luminance of each interreflecting surface is calculated by the method described in chapter 2. In this case, however, the finite element equations are solved three times, once for each primary colour, resulting in three final luminance values, l_R', l_G' and l_B', for each surface. These numbers, after gamma correction (see section 3.52) and voltage conversion, define the video signal applied to the television display.

Since the interreflection equations are solved only three times compared to the forty required by the multi-band approach of section 3.61 calculation time is much reduced (< 3 seconds compared to 1 minute). The data input to define the surface colours is also greatly reduced, the forty reflectance values of the multi-band model being replaced by three numbers, u_s, v_s and ρ_s, for each surface. In a microprocessor system in which storage is at a premium, this may be a significant advantage.

3.63 A co-ordinated source of surface colours.

Attempts have been made to promote a co-ordinated use of colour in the hope of simplifying colour scheme preparation. British Standard BS 4800 provides a catalogue of 88 paint colours specified 'To bring the colours of the various materials used in building into systematic relationship with each other and to combine economy in the total number of colours used with sufficient flexibility for meeting technical and design requirements'. A colorimetric specification of these colours is given in BS 4800 Supplement 1 for illuminant D_{65}. Colours designed to match the chromaticities defined in BS 4800 when illuminated by D_{65} are now produced by several paint manufacturers, providing the system user with a convenient visual reference for selection of surface colour.

The three equivalent reflectance approach may be extended to include the colour rendering effects of other light sources, but clearly in this case the chromaticities of BS 4800 (Supplement 1) are no longer useful. The chromaticity and reflectance of each surface colour for each light source would first have to be defined and the normalised white point for the display adjusted to match the chromaticity of this light source. A separate transform matrix for the calculation of ρ_R, ρ_G and ρ_B is therefore required for each light source. Calculation of the chromaticities matching a particular surface colour is a laborious process, as demonstrated in section 3.61, a more direct, albeit less accurate method of chromaticity transformation is described in section 3.65.

3.64 A split-flux method for colour interreflection.

The approach used in section 3.62 for introducing colour to interreflection may also be applied to the split-flux method discussed in chapter 2. The method, although less accurate where interreflection predominates, gives a direct solution without the need to solve simultaneous equations. For computers giving a calculation time greater than three seconds using the method of section 3.62, this approximation may be preferred, particularly where the system is required to be interactive.

3.65 A direct method for chromaticity prediction under various light sources.

Sections 3.61, 3.62 and 3.64 describe techniques which allow colour to be included when using the interreflection models of chapter 2. The multi-band approach of section 3.61 allows, in principle, for the use of a light source with any spectral power distribution and surface spectral reflectance curves of any chosen shape. The colour rendering properties of the light source are therefore implicit in the calculation. The disadvantage of this approach concerns the long calculation and hence program run time and the relatively large amount of storage space necessary for the data describing light sources and surface reflectances.

The latter disadvantages are partially overcome by the method of section 3.62, but, unfortunately, for the study of colour rendering, the chromaticity and reflectance of each surface colour must be known for the light source being used. It is not possible to account for the colour rendering of a different light source by merely transforming its colour point. Thus, although providing rapid results, the method as described in section 3.62 is restricted to use of one light source.

In a technique suggested by Lynes (1974)[6], the colour rendering properties of an illuminant are described by the shape and position in uniform colour space of an ellipse. A hue circle is first generated by selecting colour samples representative of real pigments (in this case Munsell samples with chromaticities shown in Figure 3.28 were used). Twenty of these samples, of chroma 6 and value 6, have their chromaticities adjusted slightly so as to be equidistant from the D_{65} chromaticity point. Thus, when illuminated by illuminant D_{65} the modified samples will have chromaticities lying on the circle shown in Figure 3.29.

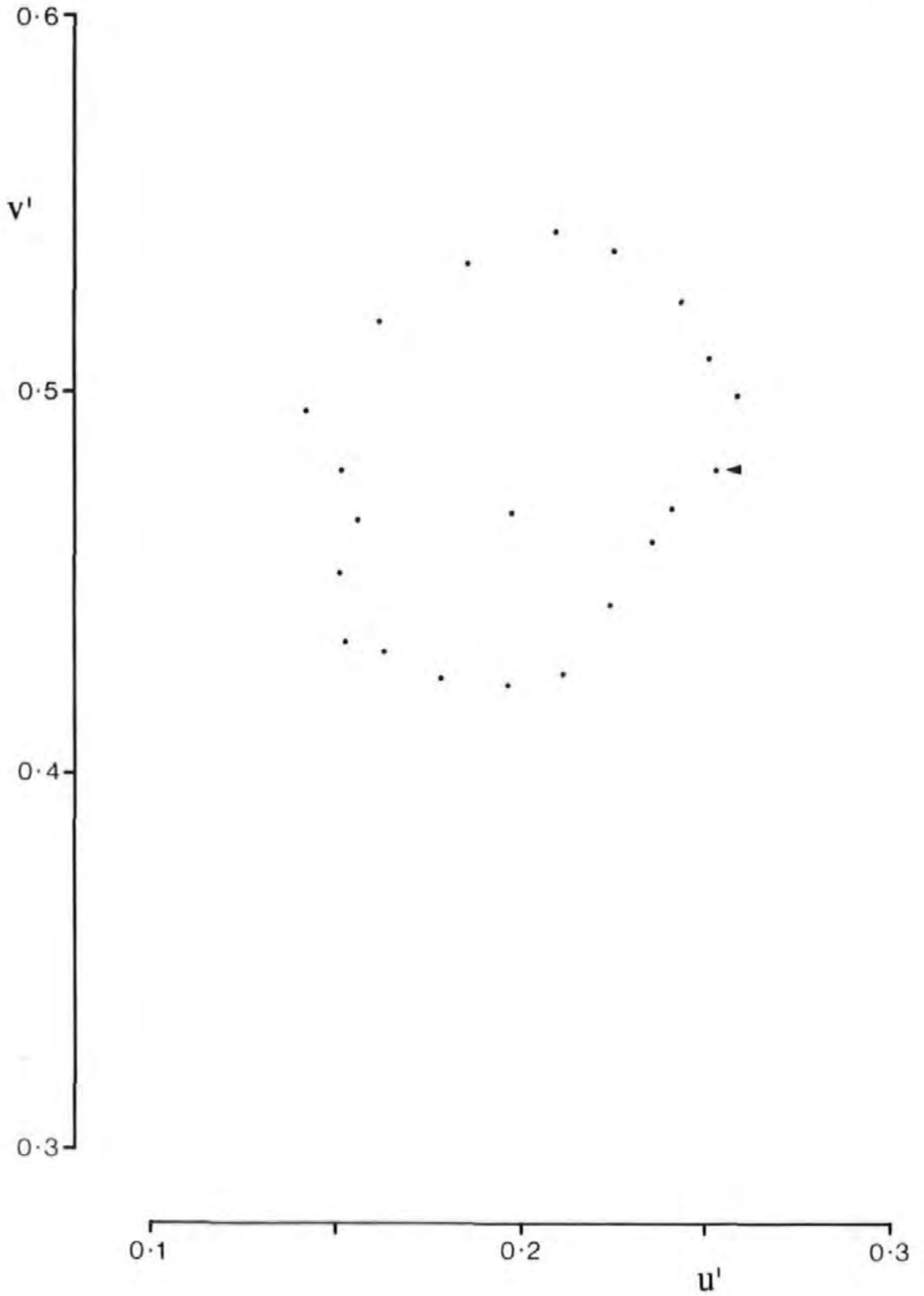


Fig. 3.28 Chromaticity co-ordinates for twenty Munsell samples with light source illuminant D_{65}.

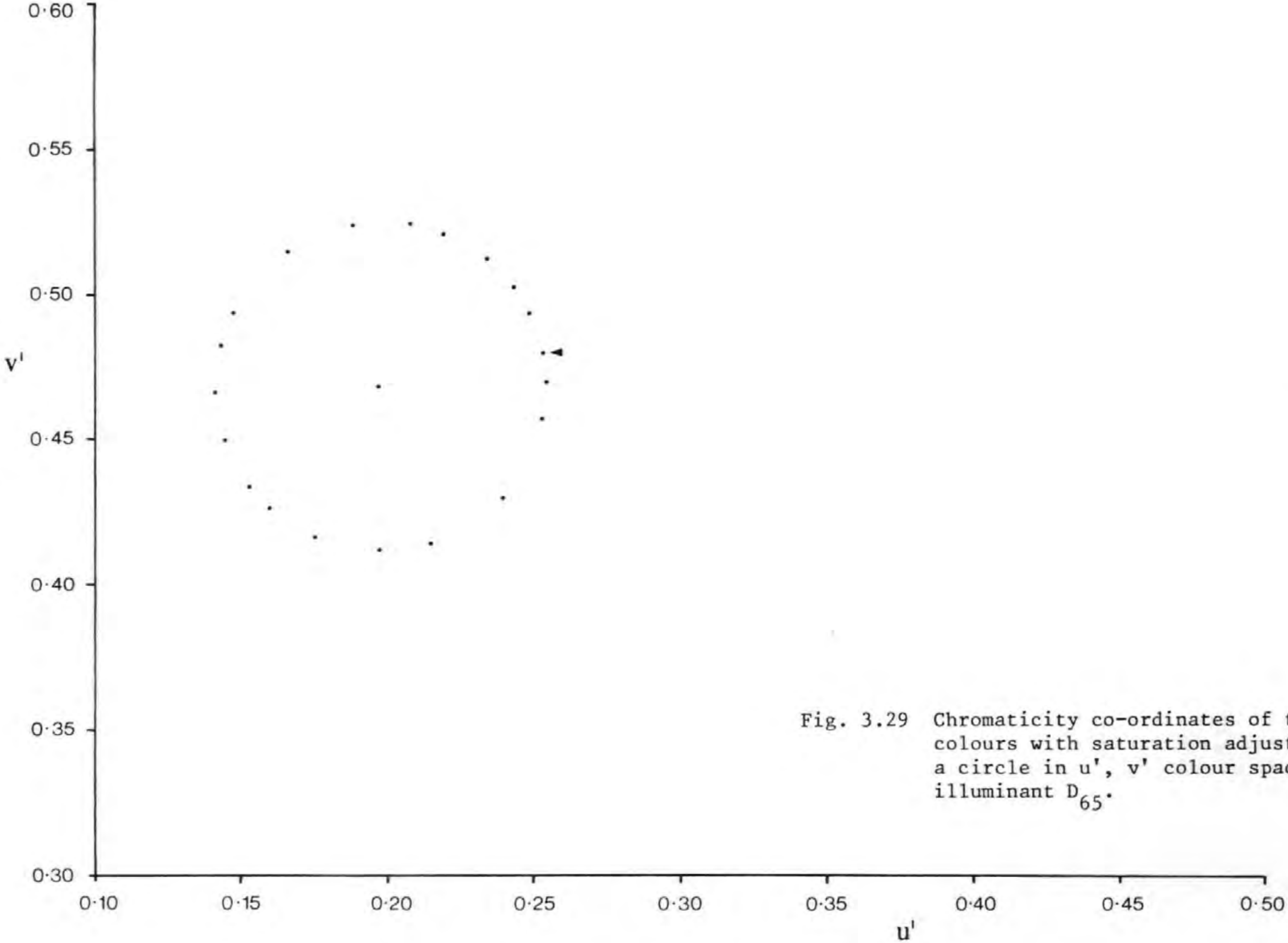


Fig. 3.29 Chromaticity co-ordinates of twenty test colours with saturation adjusted to form a circle in u', v' colour space with illuminant D_{65}.

When illuminated by other illuminants, such as an incandescent or fluorescent lamp, the modified samples have chromaticities which form ellipses, the centre of which is close to the chromaticity point of the particular light source as illustrated by Figures 3.30 to 3.40. Using a simple mathematical transformation it is now possible, in principle, to predict the chromaticity, under various illuminants, of a sample which in average daylight has a chromaticity lying on the D_{65} hue circle.

The transformation procedure is illustrated by Figure 3.41. The hue angle θ is first determined for the surface chromaticity in average daylight. For another light source, in this case illuminant A, the surface chromaticity can now be predicted by first fixing the point P on the auxiliary circle of the relevant ellipse. The required chromaticity (point T) is, to a good approximation, the point where the perpendicular PN meets the ellipse.

In this form, the transformation is only useful for daylight chromaticities lying on the D_{65} hue circle. To account for chromaticities not on this circle, a further transformation is necessary. Given a surface which in daylight has a chromaticity u,v as shown in Figure 3.42 by mixing light directly from the illuminant (u_1,v_1) with light reflected from the surface, the resultant mixture will have a chromaticity lying on the straight line joining u_1,v_1 with u,v. The position of the resultant chromaticity is then dependent only on the relative luminances of the light source L_1 and the surface L_2. Thus the amount of 'white' light required to be added to the chromaticity u,v to produce a resultant colour point u_M,v_M, lying on the D_{65} hue circle, is found. The ellipse transformation is now carried out as described above, giving the point u_{MT},v_{MT}. Finally, the transformed chromaticity u_T,v_T is found using the assumption that the ratio L_1/L_2 is the same for the ellipse as for the circle, by again applying the formula for colour mixture. The transformation assumes that surface reflectance is unaffected by the changes of illuminant.

Computer programs based on the models described in this chapter are listed in appendix 3C.

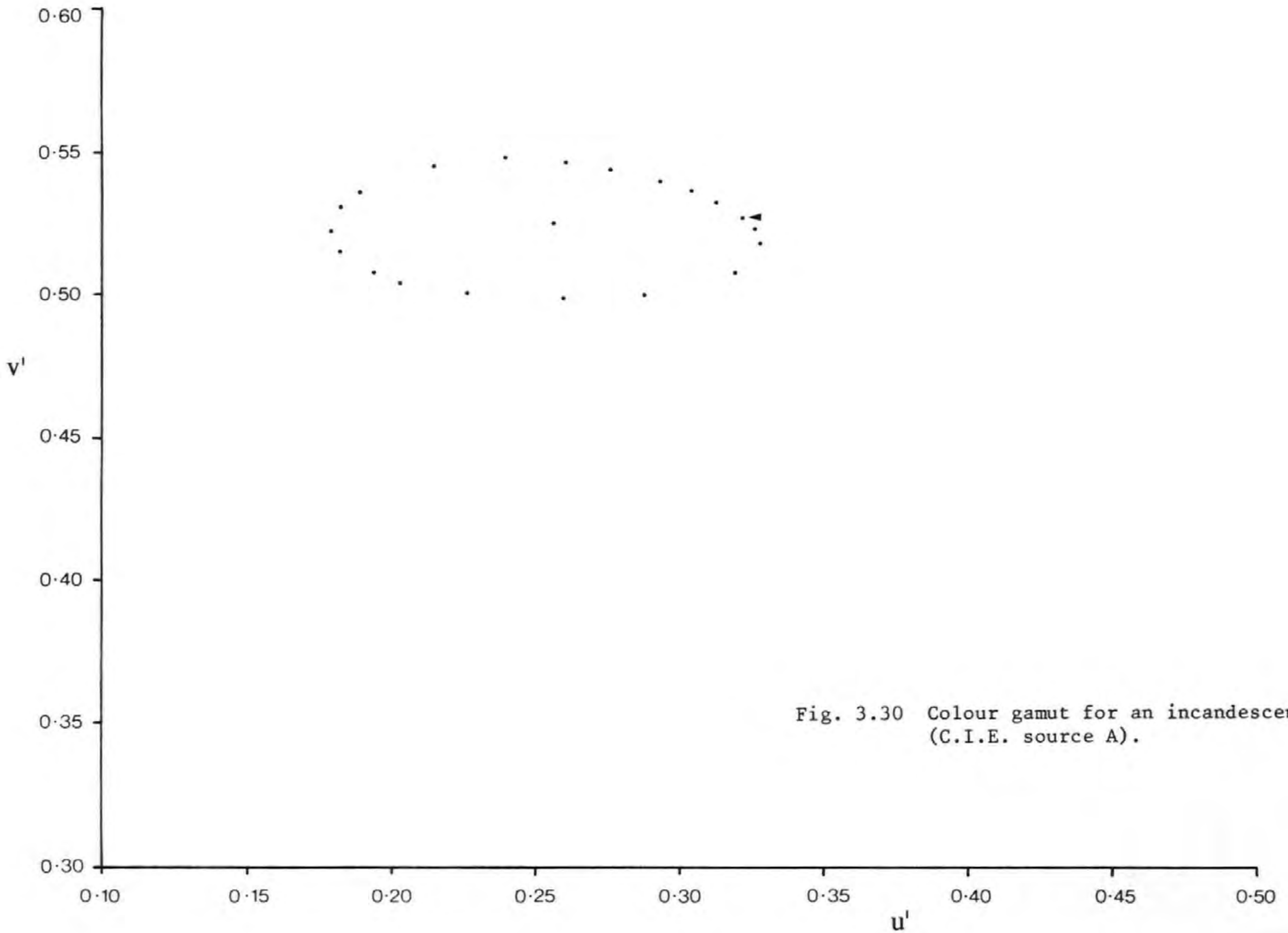


Fig. 3.30 Colour gamut for an incandescent lamp (C.I.E. source A).

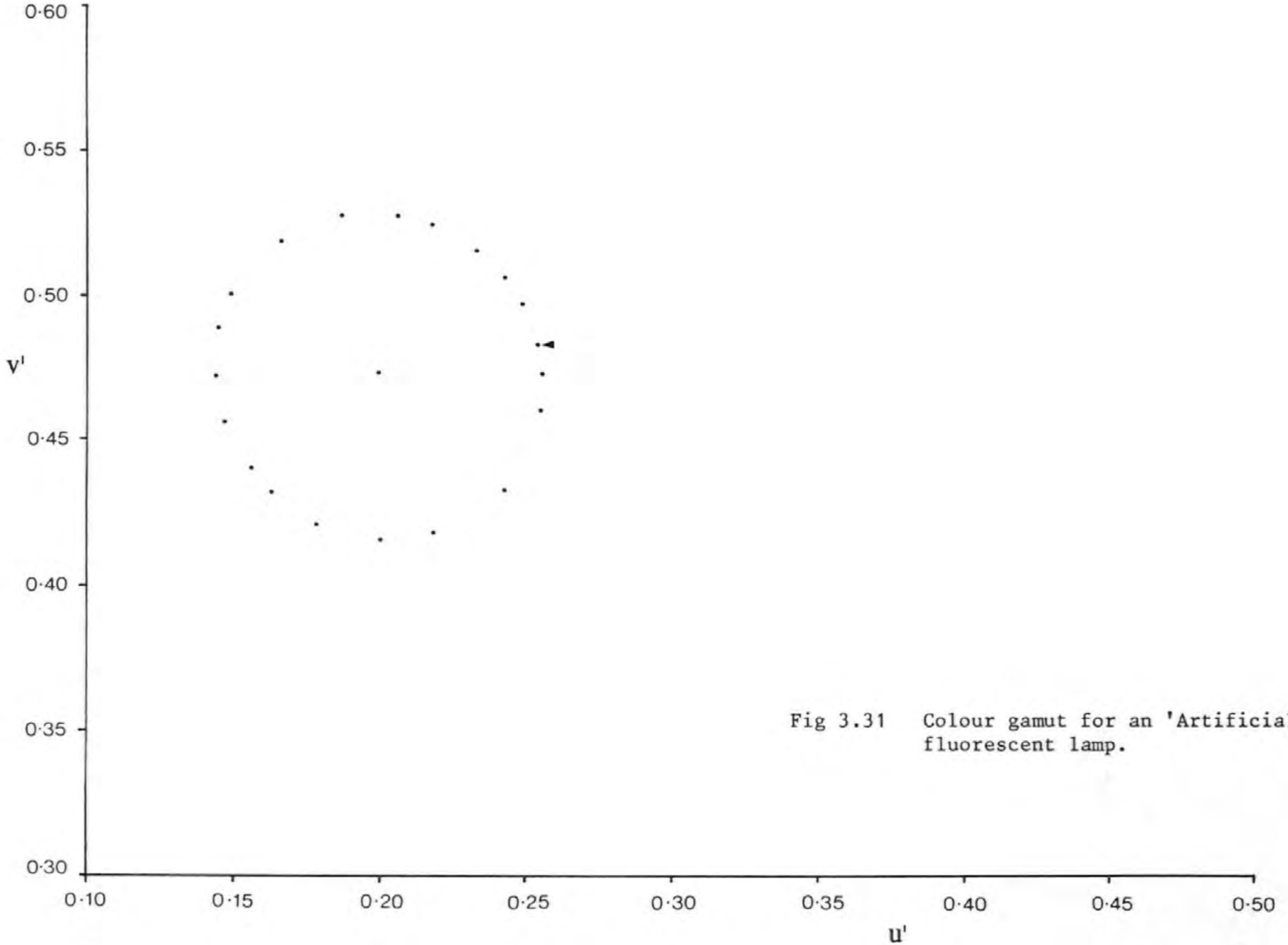


Fig 3.31 Colour gamut for an 'Artificial Daylight' fluorescent lamp.

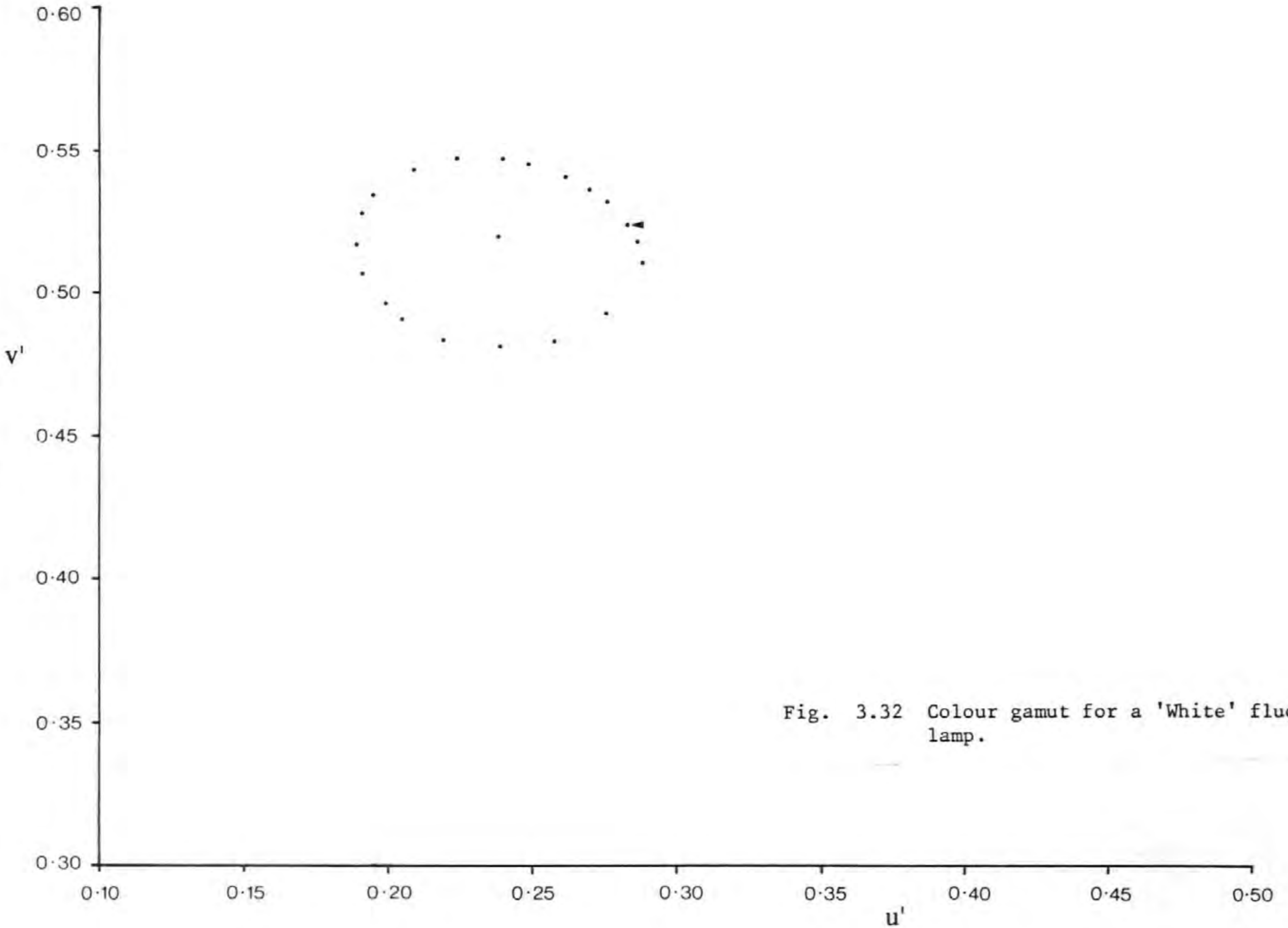


Fig. 3.32 Colour gamut for a 'White' fluorescent lamp.

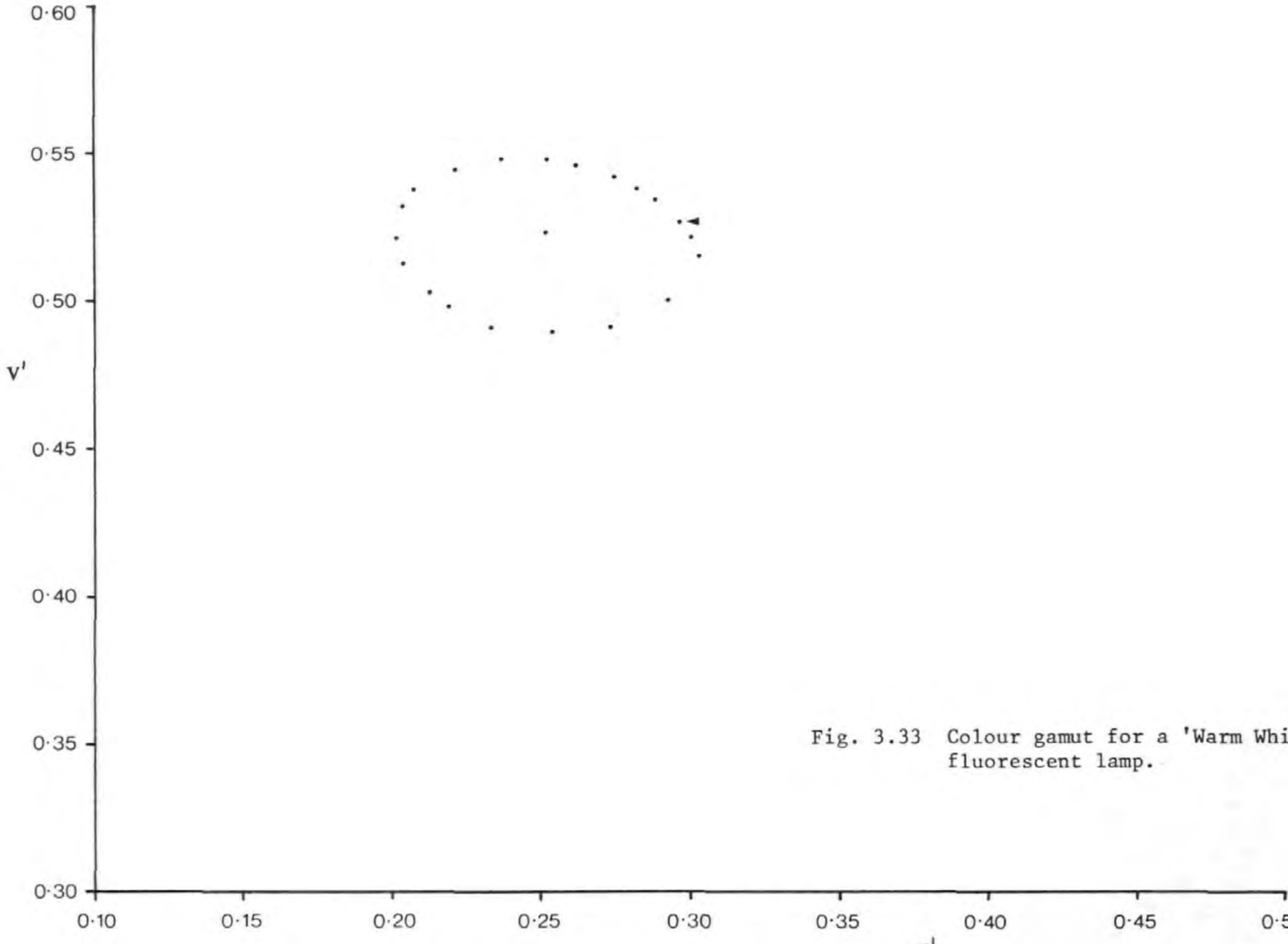


Fig. 3.33 Colour gamut for a 'Warm White' fluorescent lamp.

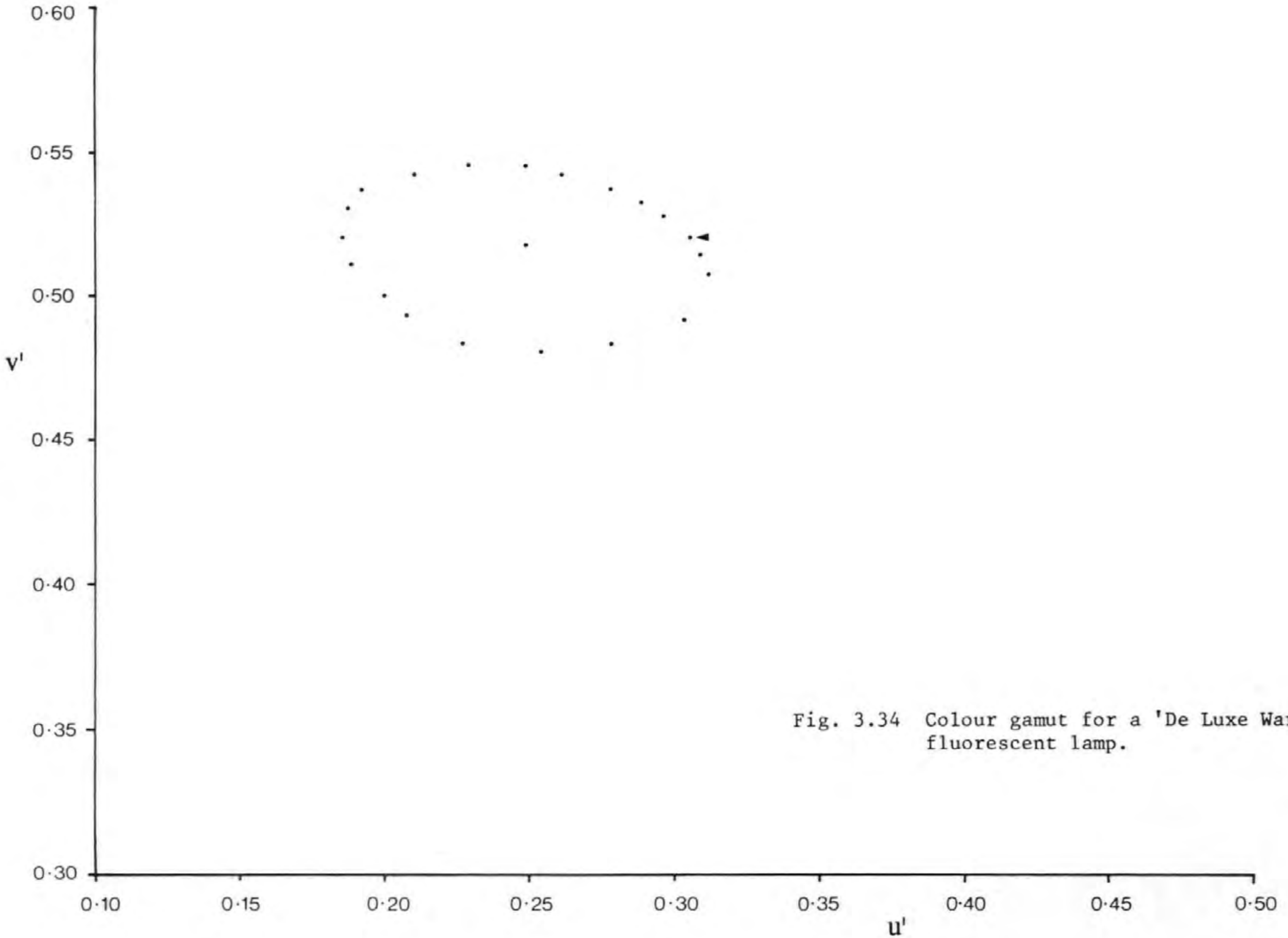


Fig. 3.34 Colour gamut for a 'De Luxe Warm White' fluorescent lamp.

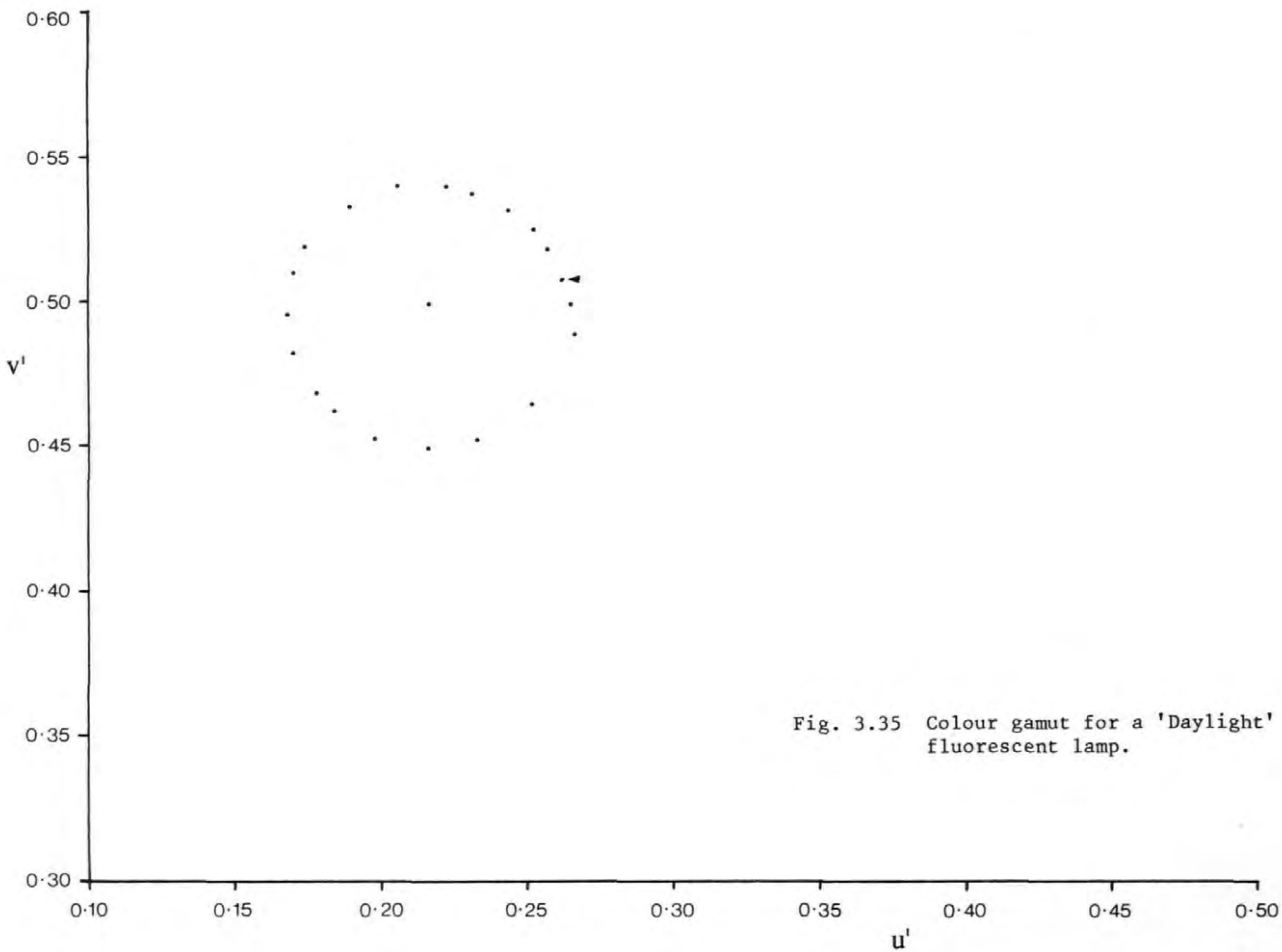


Fig. 3.35 Colour gamut for a 'Daylight' fluorescent lamp.

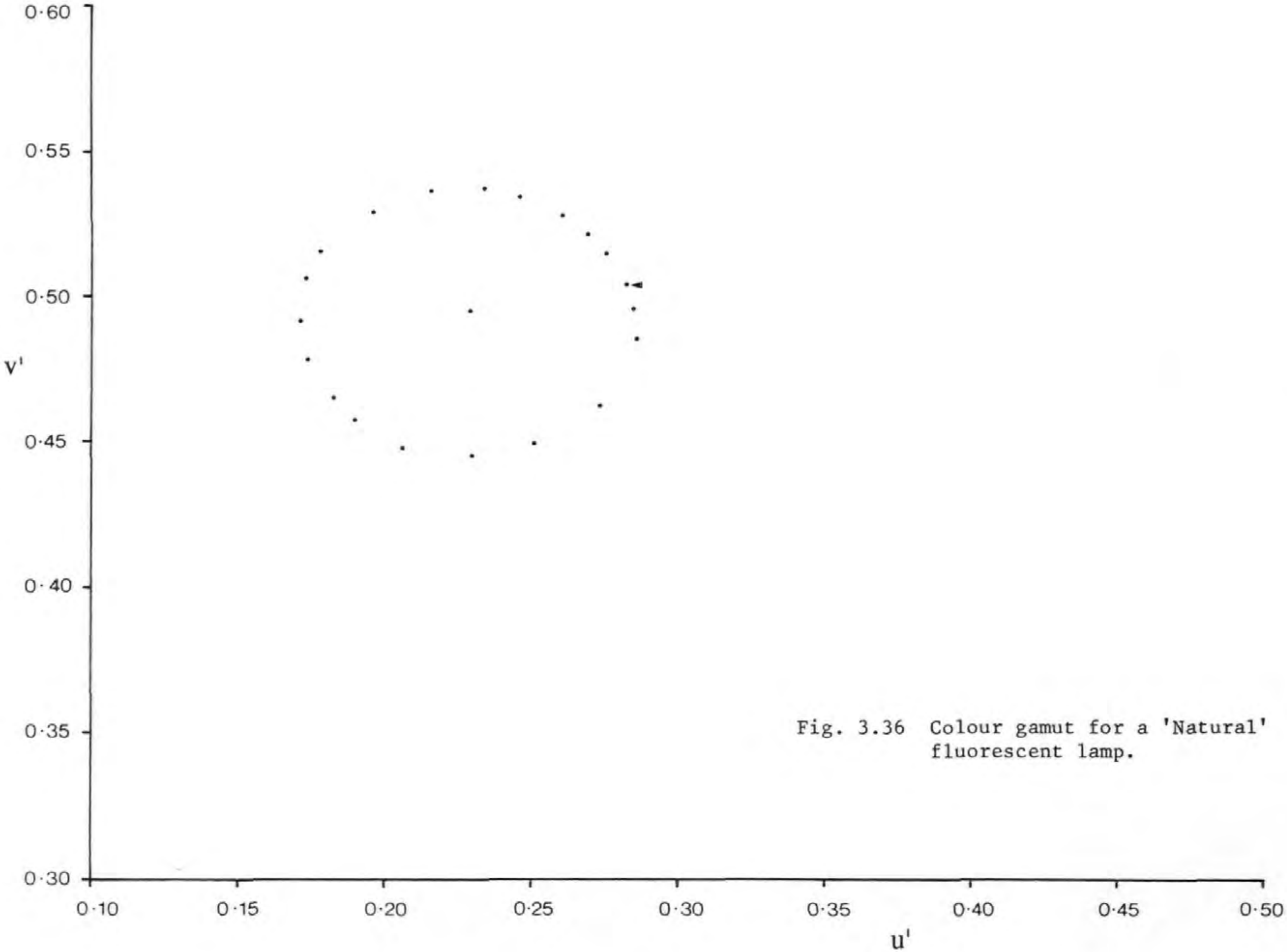


Fig. 3.36 Colour gamut for a 'Natural' fluorescent lamp.

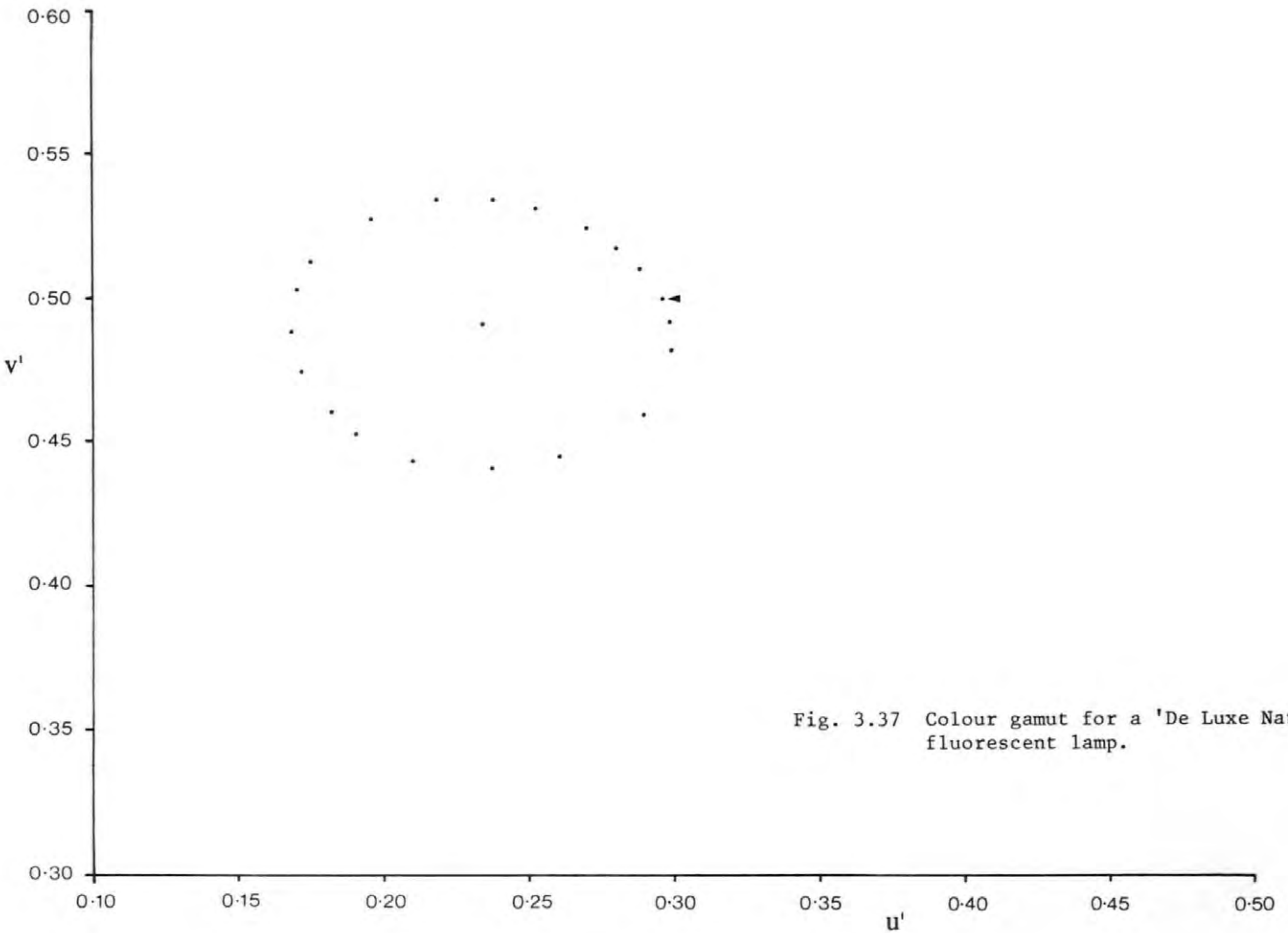


Fig. 3.37 Colour gamut for a 'De Luxe Natural' fluorescent lamp.

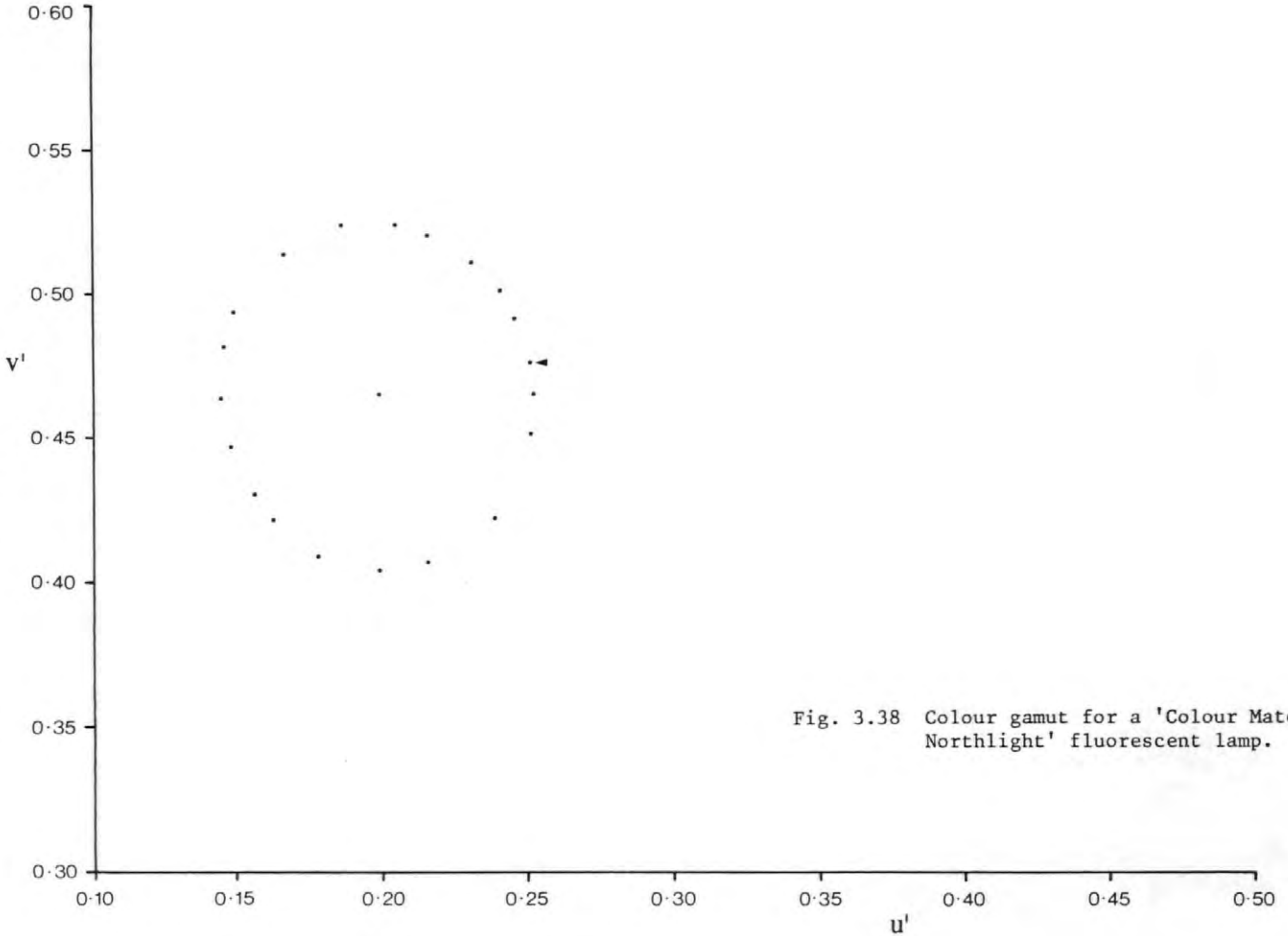


Fig. 3.38 Colour gamut for a 'Colour Matching/ Northlight' fluorescent lamp.

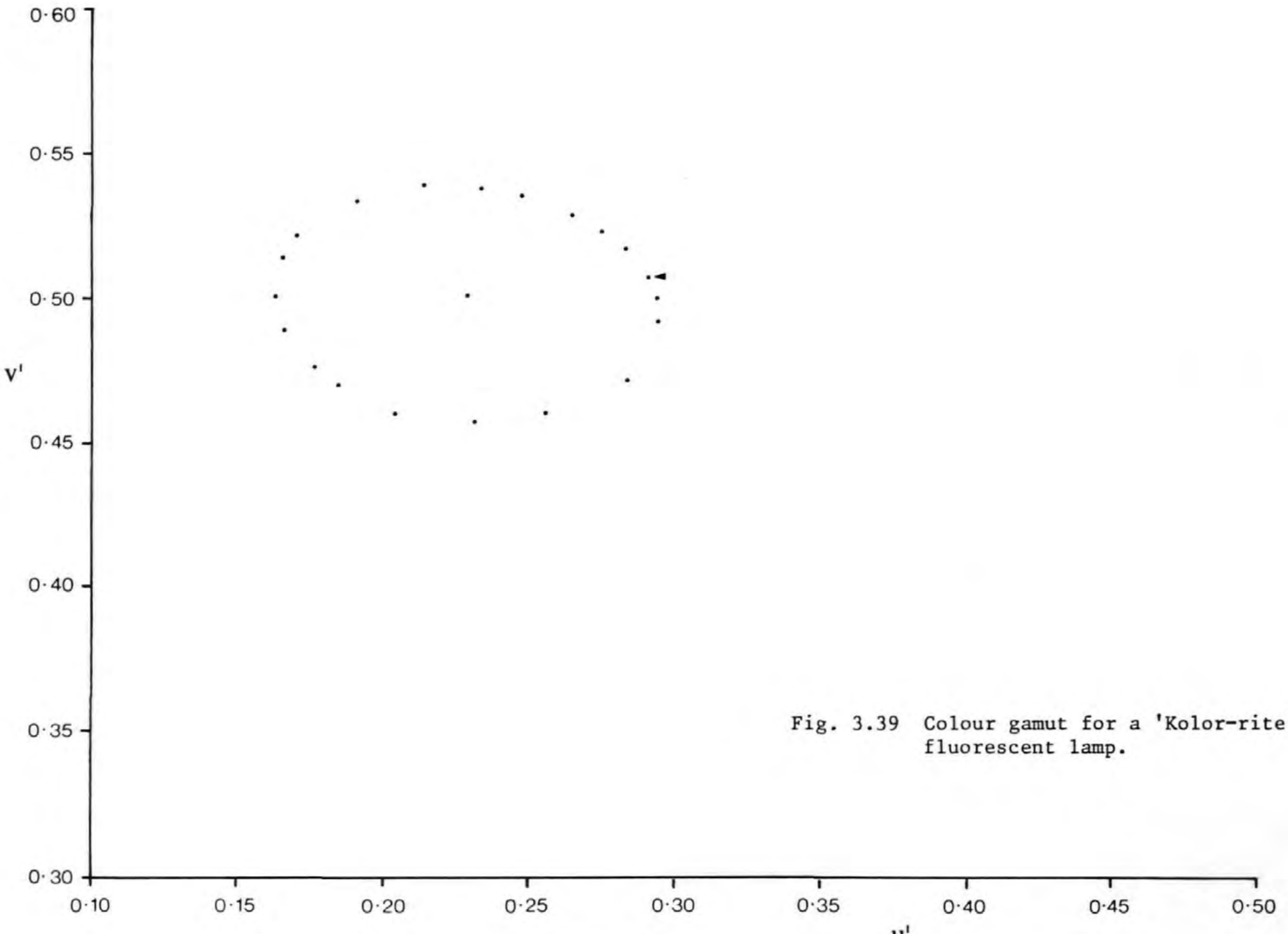


Fig. 3.39 Colour gamut for a 'Kolor-rite' fluorescent lamp.

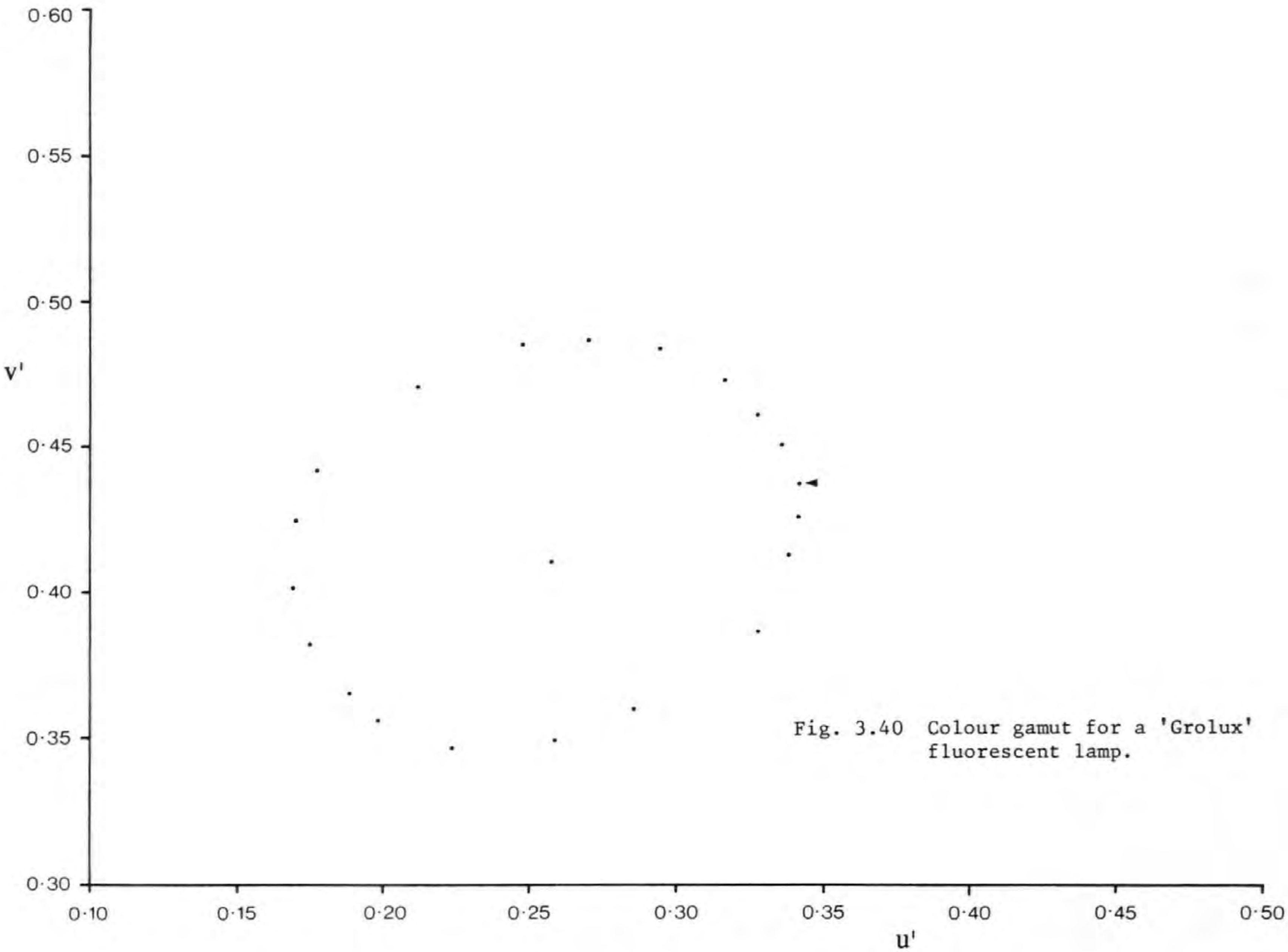


Fig. 3.40 Colour gamut for a 'Grolux' fluorescent lamp.

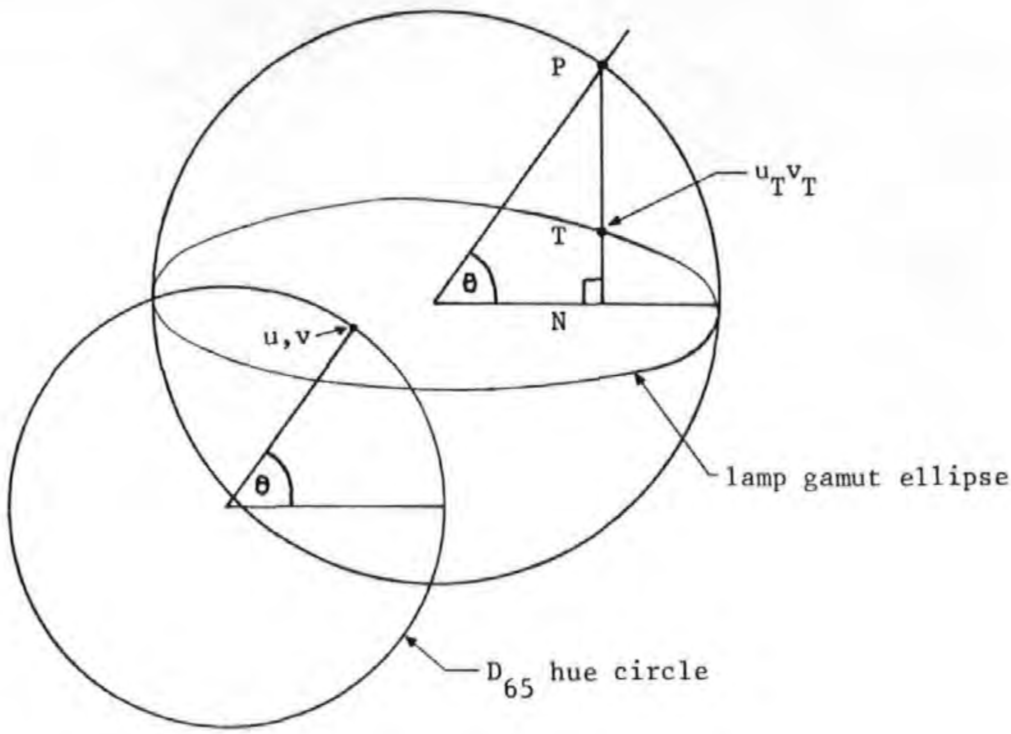


Fig. 3.41 The basic colour transformation.

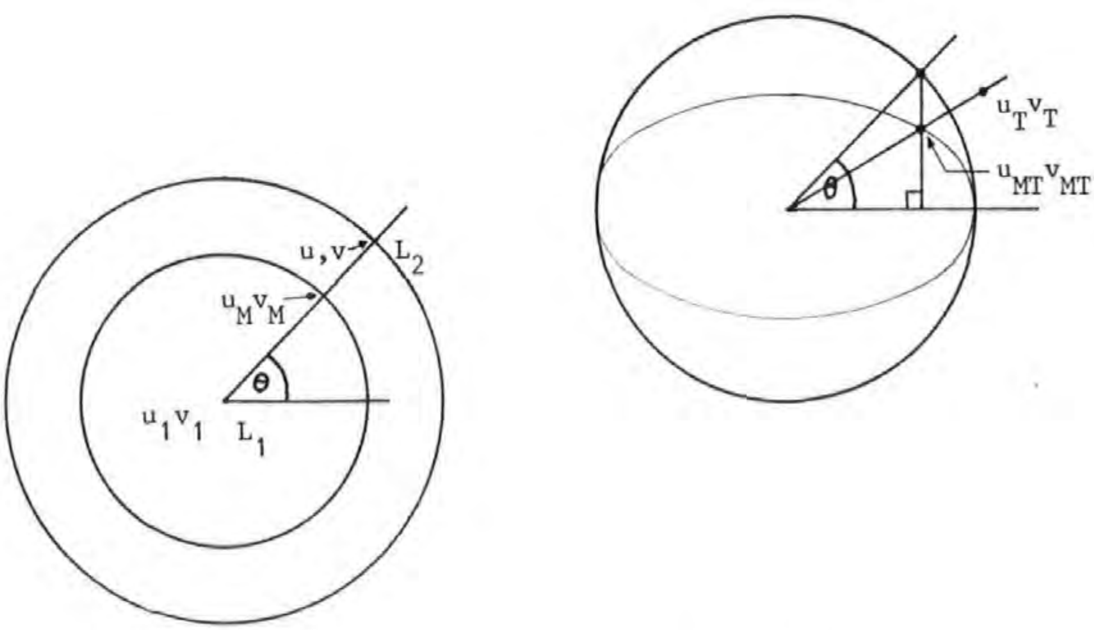


Fig. 3.42 Transformation for colours not lying on the hue circle.

References

1. Lynes, J.
   Interreflection and Flux Distribution in Lighted Interiors.
   I.E.S. Monograph no. 2, p. 7 (1957).

2. Guild, J.
   Phil. Trans. Roy. Soc. (London), A, 230, p. 149 (1931).

3. Wright, W. D.
   Trans. Opt. Soc., vol. 30, p. 141, (1928-29).

4. MacAdam, D. L.
   J. Opt. Soc. Am., vol. 27, p. 294 (1937).

5. C.I.E. Colorimetry Committee.
   J. Opt. Soc. Am., vol. 64, p. 896 (1974).

6. Lynes, J.
   Lamps for Hue Discrimination - A Preliminary Study. Internal Report no. 10/74, Plymouth Polytechnic School of Architecture (1974).

7. The Munsell Color Company.
   Munsell Book of Color, vol. 1 (1929).

8. Newhall, M., Nickerson, D. and Judd, D. B.
   J. Opt. Soc. Am., vol. 33, p. 385 (1943).

9. Keegan, H. J. et al.
   J. Opt. Soc. Am., vol. 48, p. 863 (1958).

10. Reinboldt, W. C. and Menard, J. P.
    J. Opt. Soc. Am., vol. 50, no. 8, p. 802 (1960).

11. Wyzecki, G.
    J. Opt. Soc. Am., vol. 53, p. 1318 (1963).

12. Bartleson, C. J. and Breneman, E. J.
    J. Opt. Soc. Am., vol. 57, p. 953 (1967).

13. Booth, J. M. and Schroeder, R. B.
    Computer, vol. 10, 8, p. 15 (1977).

14. Joblove, G. H. and Greenberg, D.
    Colour Spaces for Computer Graphics. Computer Graphics, vol. 12, p. 20 (1978).

15. Smith, A. R.
    Computer Graphics, vol. 12, p. 12 (1978).

16. Sloan, K. R. and Brown, C. M.
    Computer Graphics and Image Processing, vol. 10, 4, p. 317 (1979).

17. Wood, C. B. B. and Sproson, W. N.
B.B.C. Engineering, vol. 19, p. 36 (1971).

18. Hunt, R. W. G.
J. Opt. Soc. Am., vol. 55, p. 1540 (1965).

19. Bartleson, C. J. and Witzel, R. F.
Phot. Sci. Eng., vol. 11, p. 329 (1967).

20. B.B.C. Research Department Report no. 1971/4 (1971).

21. Judd, D. B. and Wyszecki, G.
Color in Business, Science and Industry. J. Wiley and Sons, New York (1963).

22. MacAdam, D. L.
J. Opt. Soc. Am., vol. 32, p. 247 (1942).

23. Wright, W. D.
Proc. Phys. Soc. (London), vol. 53, p. 93 (1943).

24. Judd, D. B.
J. Opt. Soc. Am., vol. 26, p. 421 (1936).

25. Pearson, D. E.
The Transmission and Display of Visual Information. p. 164, Pentech Press (1975).

26. International Electrotechnical Commission. Photometric and Colorimetric Methods of Measurement of the Light emitted by a Cathode-ray Tube Screen.
Publication 441 (1974).

27. Wyszecki, G. and Stiles, W. S.
Color Science. p. 16, Wiley and Sons (1967).

28. Philippart, H. A. S.
Tristimulus Spot Colorimeter. B.B.C. Engineering Monograph no. 65 (1966).

29. Hacking, K.
A tristimulus colorimeter for television. 122nd meeting of the Colour Group, (GB), City University, London, November (1976).

Chapter 4

HARDWARE DESIGN

4.1 Introduction

This chapter describes some of the problems encountered in generating simple patterns on a raster scan display. A particular display requirement is presented along with a possible approach to the hardware design problem.

A perspective representation of a room interior is simply defined by a geometry shown in Fig 4.01. The most popular method of generating such an image uses a digital memory or 'frame store' from which data is output serially at the video scan rate, each store location being dedicated to one pixel in the final image. For a high quality grey-scale picture, each store location requires typically 7 bits, the total store requirement being at least 64K x 7 bits. The large video memory requirement means that such a system, with the addition of colour, would cost approximately £10,000 for the frame store alone.

For the relatively simple graphic image presentation of Fig 4.01 it is possible to avoid the use of a frame store; costs are therefore much lower. The image is produced by a dedicated hardware system which can, in principle, be designed using analogue or digital circuitry. Using analogue methods, variables are represented in continuous form, their accuracy being limited by the quality and hence cost of the components used.

A maximum accuracy of 0.01% of maximum range would be typical

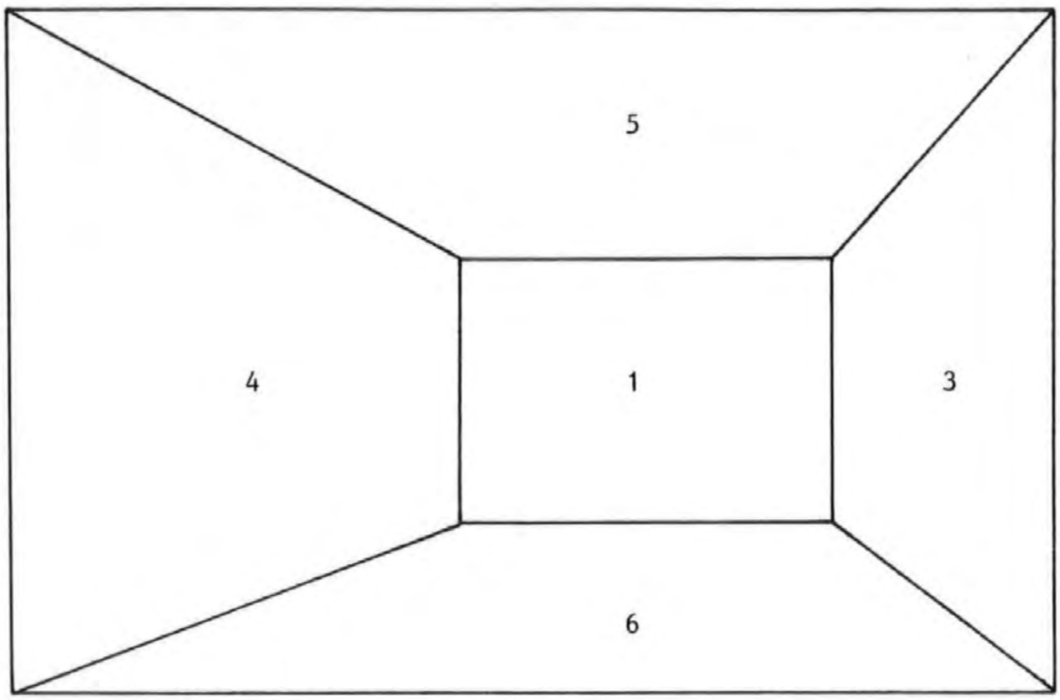


Fig. 4.01 The geometry of the final image.

for each variable.

If variables are represented as discrete but precisely defined steps as is done in digital systems, then there is in principle no limit to the accuracy which may be attained, this being related simply to the number of bits used to represent the variables. The recent growth in demand for digital circuit components, largely by the computer industry, has led to considerable cost reduction, making digital systems preferable, in many cases, on grounds of costs alone.

The refresh requirement of a raster-scan display makes necessary the storage of display variables. Both digital and analogue approaches can provide storage. However, analogue storage is expensive and can give rise to inaccuracy since the electronic charge representing a variable will leak away with time. Analogue circuits usually depend on the accurate and stable setting of variable resistances which are sensitive to changes of temperature and mechanical vibration. In contrast, digital numbers may be stored indefinitely with no loss of accuracy. Digital systems do not, in general, require the tedious 'setting up' and calibration so often a characteristic of analogue circuitry. Thus, the long term stability and reproducibility of digital circuits are superior, in most cases, to their analogue equivalents.

In view of the advantages of digital circuitry as outlined above, the display system described in this chapter is based on digital techniques wherever possible. A diagram of the major system components is given in Fig 4.02. In cases where analogue devices have to be used, such as the video input amplifier of the colour television monitor, digital to analogue conversion is necessary to

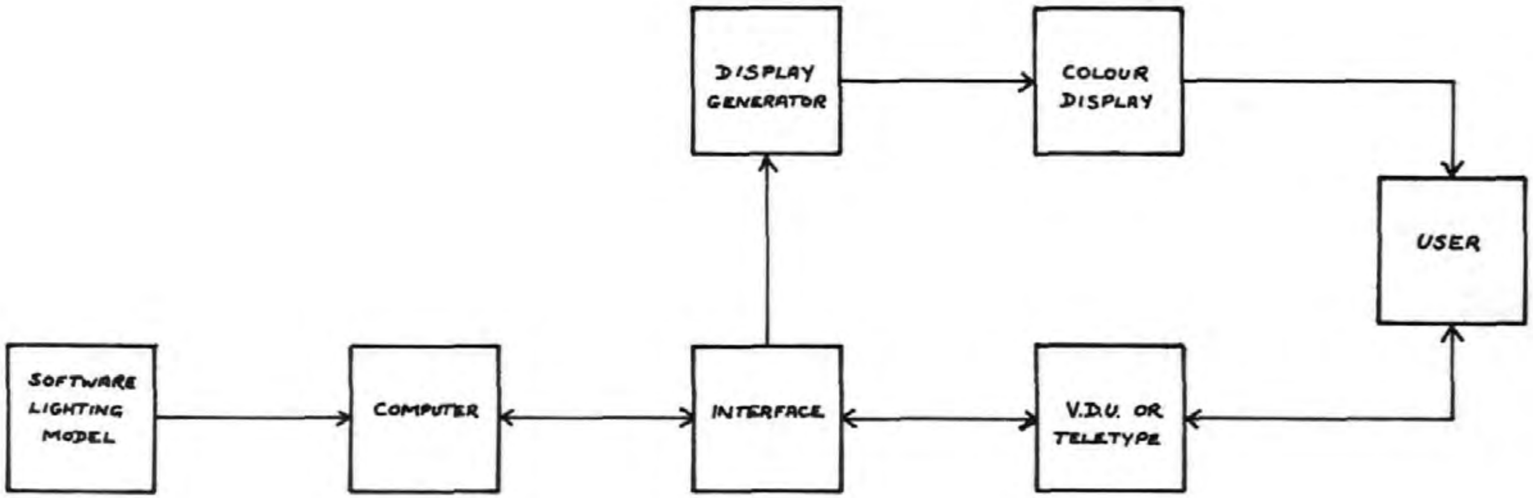


Fig. 4.02 The major system components.

produce a compatible signal.

4.2 Generation of a line segment using a raster-scan display

The European standard for broadcast television provides 575 active display lines arranged in a 2:1 interlace as in Fig 4.03 This form of scan is usually produced by magnetically deflecting an electron beam in two perpendicular directions. Conventionally the vertical scan rate (50 Hz) is much slower than the horizontal rate (15.625 kHz). The resulting picture is an assembly of horizontal lines. In the vertical direction, therefore, the beam position is discrete whilst horizontally the position is, at least in principle, continuous. In practice, the horizontal resolution is determined by the frequency response of the video amplifiers.

Interlacing is a technique employed by the television system designers to minimise the bandwidth requirement for television transmission whilst avoiding subjective flicker. A complete frame of video information is displayed in two parts. Firstly only the alternate (even numbered) lines are scanned, on completion of which the electron beam retraces to the top of the frame and commences scanning the odd numbered lines. The two interlaced scans are referred to respectively as odd and even fields, as shown in Figs 4.03 and 4.08. The field rate is 50 Hz but the frame rate is effectively reduced to 25 Hz. Interlacing allows this low frame rate to be employed by merging spatial and temporal effects such that a flicker-free image is perceived.

With a raster display, a line is represented by a series of bright-up points on successive scan lines. Since the vertical position is implicitly discrete, the y co-ordinate may be defined

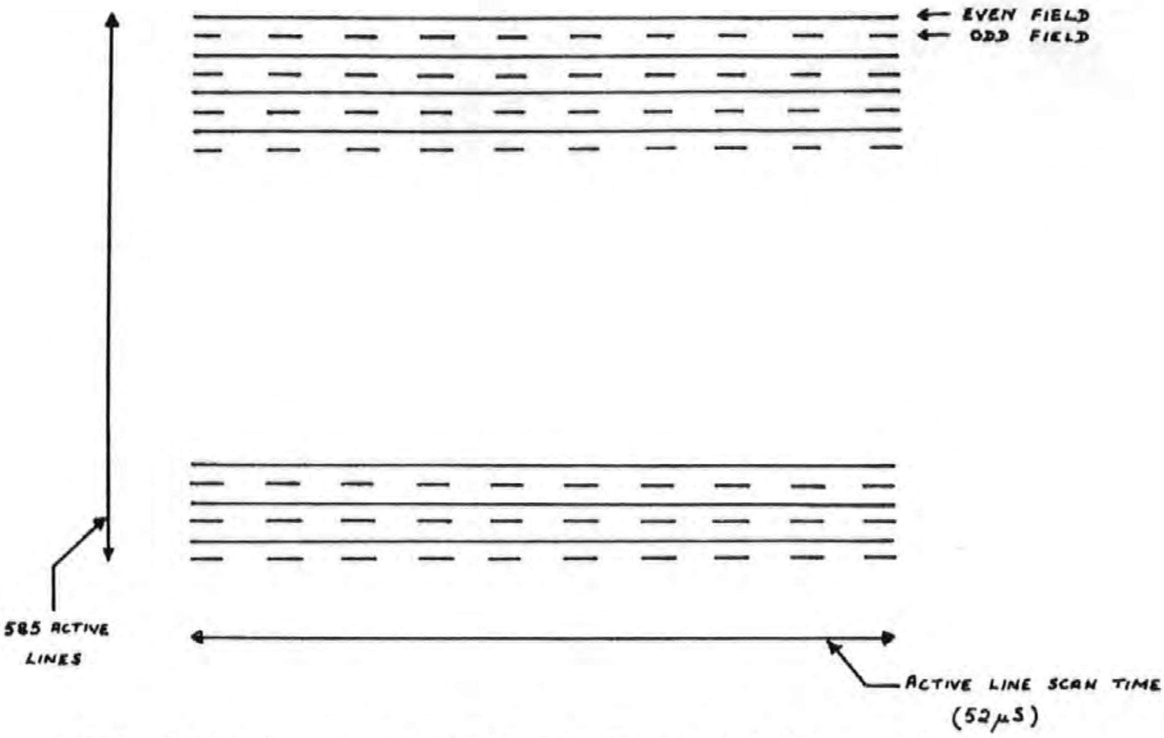


Fig. 4.03 Raster format for the European PAL system I.

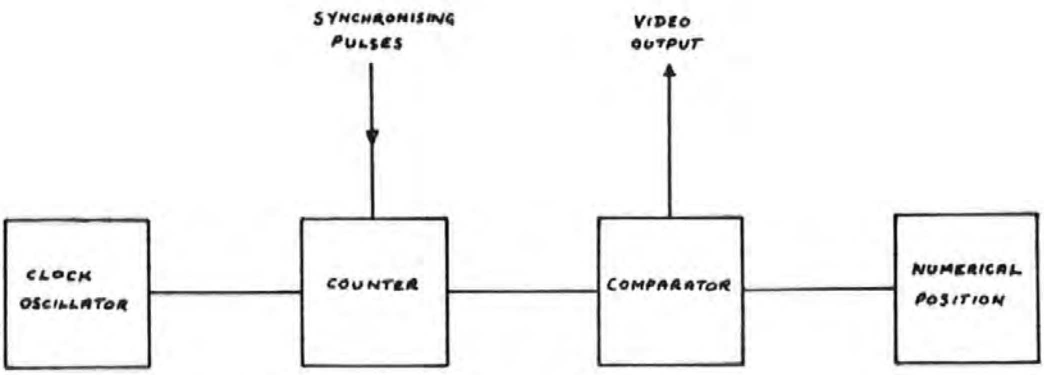


Fig. 4.04 Basic circuit for image positioning.

for a digital system by simply counting lines, using, for example, the counter-comparator arrangement shown in Fig 4.04. Digital positioning, however, places a further constraint on horizontal resolution.

A vertical discrimination of 1 in 575 requires, in digital terms, a count up to 10 bits. For the standard picture aspect ratio of 4:3, the equivalent horizontal resolution is achieved by a count of 575 x 4/3. This resolution requires a clock frequency in excess of 15 MHz. Using two counters, one incrementing at 15.625 kHz and the other at a much higher frequency of 15 MHz, with two digital comparators, the position of any point on the screen is defined by the comparator inputs x and y.

To generate a line it is necessary to define the end-points and the spatial relationship between points on successive scan lines. The simplest case of a straight line and a non-interlaced scan is shown in Fig 4.05. The line has start point A and end point B, with the location of the bright-up changing from line to line by an amount δx where

$$\delta x = \tan \phi = \frac{c - a}{d - b}$$

If j = y - b + 1, then the x co-ordinate required on line y = b + j - 1 is given by

$$x_{b+j-1} = a + (j-1)\delta x \quad \text{for} \quad j = 1, 2, \cdots d-b+1$$

taking successive lines

$$\therefore \quad x_{b+j-1} = x_{b+j-2} + \delta x$$

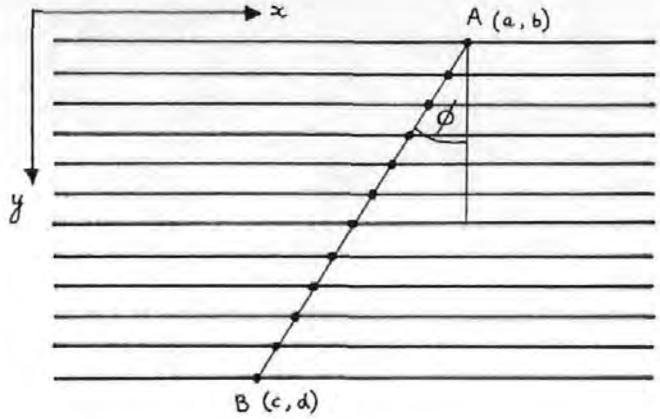


Fig. 4.05 Generation of a line segment with non-interlaced scan.

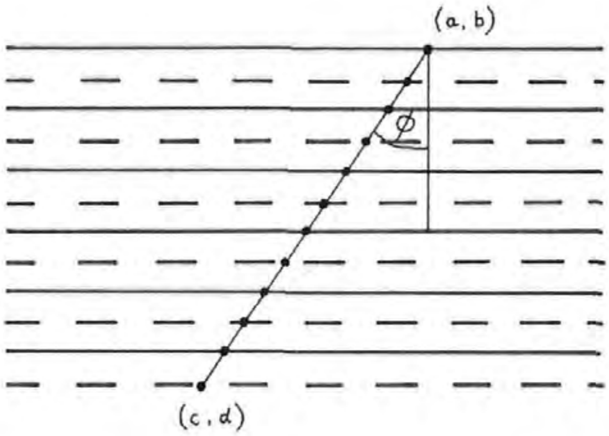


Fig. 4.06 Generation of a line segment for an interlaced scan.

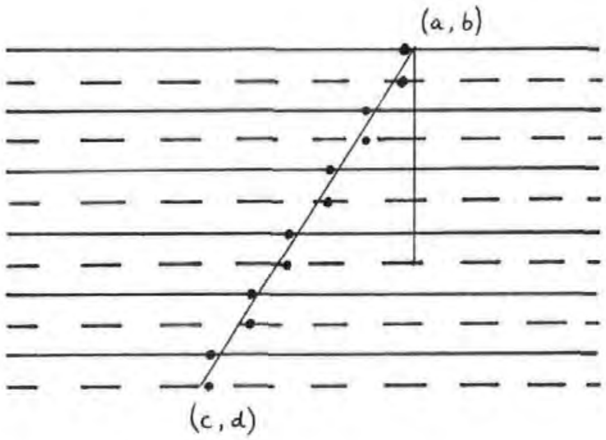


Fig. 4.07 Generation of a line segment for an interlaced scan with identical fields.

for $-90° \leqslant \phi \leqslant 0°$ $\qquad \delta x$ is -ve

for $0° \leqslant \phi \leqslant 90°$ $\qquad \delta x$ is +ve

Fig 4.06 illustrates the case for an interlaced scan. The solid lines form the odd field and the dotted lines the even field, these being scanned successively.

For the odd field, taking $j = \left[(y - b)/2\right] + 1$, the x co-ordinate required on line b + 2(j - 1) is given by

$$x_{b+2(j-1)} = a + 2(j-1)\,\delta x \quad \text{for } j = 1, 2, \cdots \left[(d-b)/2 + 1\right]$$

for successive lines

$$\therefore\ x_{b+2(j-1)} = x_{b+2(j-2)} + 2\,\delta x$$

similarly for the even field if $\bar{j} = (y - b + 1)/2$

$$\therefore\ x_{b+2\bar{j}-1} = a + (2\bar{j} - 1)\,\delta x \quad \text{for } \bar{j} = 1, 2, \cdots (d-b)/2$$

$$\therefore\ x_{b+2\bar{j}-1} = x_{b+2\bar{j}-3} + 2\,\delta x$$

for successive lines.

Finally, Fig 4.07 illustrates the case for an interlaced scan in which the information for each field is identical.

For the odd field, if $j = (y - b)/2 + 1$

$$\therefore\ x_{b+2(j-1)} = a + 2(j-1)\,\delta x \quad \text{for } j = 1, 2, \cdots \left[(d-b)/2 + 1\right]$$

for successive lines

$$\therefore\ x_{b+2(j-1)} = x_{b+2(j-2)} + 2\,\delta x$$

For the even field, if $j = (y - b + 1)/2$

$$\therefore x_{b+2j-1} = a + 2(j-1)\delta x \quad \text{for } j = 1, 2, \cdots [(d-b)/2+1]$$

for successive lines

$$\therefore x_{b+2j-1} = x_{b+2j-3} + 2\delta x$$

For all three cases outlined above a line is generated by incrementing or decrementing the x co-ordinate on successive lines. Updating of the x value is carried out only during the 12μs line blanking interval thus ensuring a stable display during active line scan. The x values may be generated by a binary counter or alternatively be produced by a read only or random access memory, as appropriate.[1]

4.21 The choice of clock frequency

The horizontal resolution is a function of the bandwidth of the video amplifiers used by the display in addition to the clock frequency used to increment the x co-ordinate. For the graphic display to be described in 4.3, a vertical resolution of 1:287 (by displaying identical information on each field) was found to be subjectively convincing. For this reason, the horizontal clock frequency was reduced from the 15 MHz discussed in 4.2 to a compromise of approximately 11 MHz.

The synchronizing pulses required to maintain the time relationship between picture information and the scan of the display monitor are generated by an L.S.I. circuit type ZNA103. This circuit produces a 50 Hz frame pulse output together with interlaced line synchronizing pulses with a period of 64μs. The output is available as a composite mixed sync signal or as separate line and frame synchronizing signals as shown in Fig 4.08. The display monitor requires a negative-going sync input signal and has a 75 ohm input impedance. The circuit necessary to match the TTL (transistor-transistor logic) type output of the ZNA103 to the 75 ohm load is shown in Fig 4.09.

The ZNA103 requires an input clock frequency of 656.25 kHz. For this reason the horizontal count frequency was chosen to be 11.8125 MHz. This frequency gives good resolution horizontally and, divided by the integer 18 using binary counters, gives the necessary ZNA103 input signal. Since all the signal sources used by the system are derived from this master oscillator, the display generator is implicitly synchronous in operation. Alternatively, synchronous operation could be maintained by use of separate signal sources in a phase-locked loop arrangement. Whilst this has the advantage of not restricting the choice of clock frequency, synchronism is dependent on the short term stability of the high frequency oscillator. The totally synchronous system, however, will maintain correct phase relationships at all times.

4.22 Hardware speed limitations.

The display circuitry is required to operate at frequencies not far from the upper limit of the TTL circuit elements from which it is developed. For this reason propagation delays and switching times can become significant. The more common TTL components are arranged in blocks of four elements. Thus a twelve bit counter or comparator is developed by cascading three, four-bit components. As more components are cascaded the propagation delay to the output increases

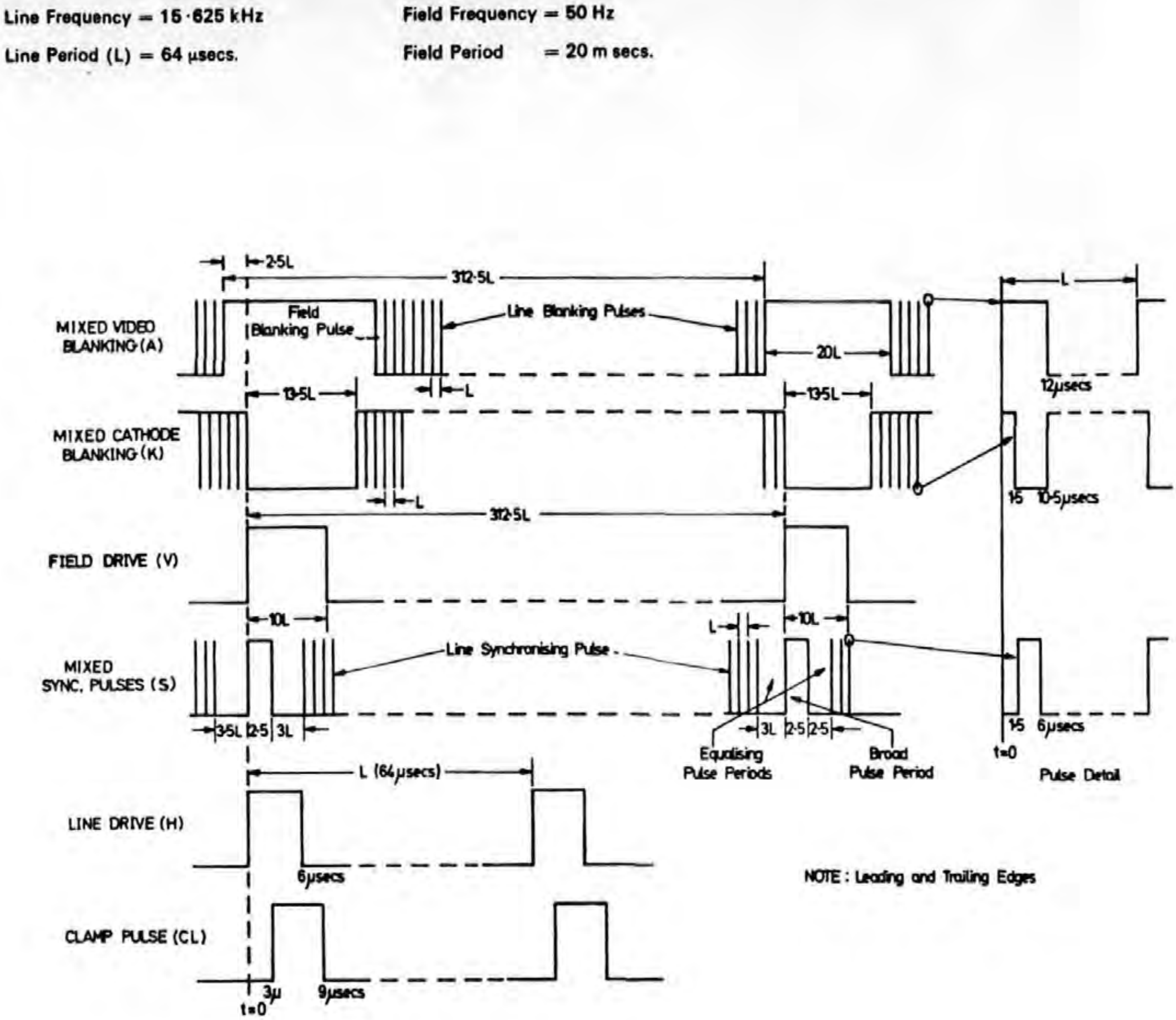


Fig. 4.08 Synchronizing waveforms used by the display system.

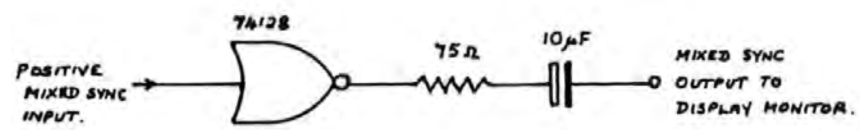


Fig. 4.09 Synchronizing output circuit for the display monitor.

in proportion and can become comparable with the clock period. The problem of propagation delay is minimised by the use, where possible, of components which operate synchronously.

4.3 Generation of a simple perspective shape.

A block diagram of the complete perspective generator is given in Fig 4.10. Four counters define the complete perspective pattern. Counters A and B are used to define horizontal and vertical position respectively. The reversible counters C and D are used to generate the interfaces 5-4, 6-4 (counter C) and 5-3, 6-3 (counter D) shown in Fig 4.01 and again in colour in Fig 4.11. These variable gradients are produced by the counter/comparator arrangement described in section 4.2 for generation of a line segment. Clock pulses are applied to C and D only during the 12μs line blanking intervals. These counters are incremented or decremented to give positive or negative gradients respectively. The line blanking pulse is used to clear counter A, whilst counter B is cleared during the field blanking interval.

The sequence of pattern generation begins at the end of the field 0 blanking interval. At the beginning of the first active display line counter C is preset to zero and counter D preset to 614_{10}. As the scan begins comparator C will give an immediate output since both A and C are set to zero. Comparator D gives its output when counter A reaches 614_{10}, at the end of the active line period. The following 12μs comprises the line blanking interval during which time counter A is cleared and counters C and D are clocked at rates dependent on the gradients 5-4 and 6-3 respectively. The horizontal count sequence begins again at the start of the second line

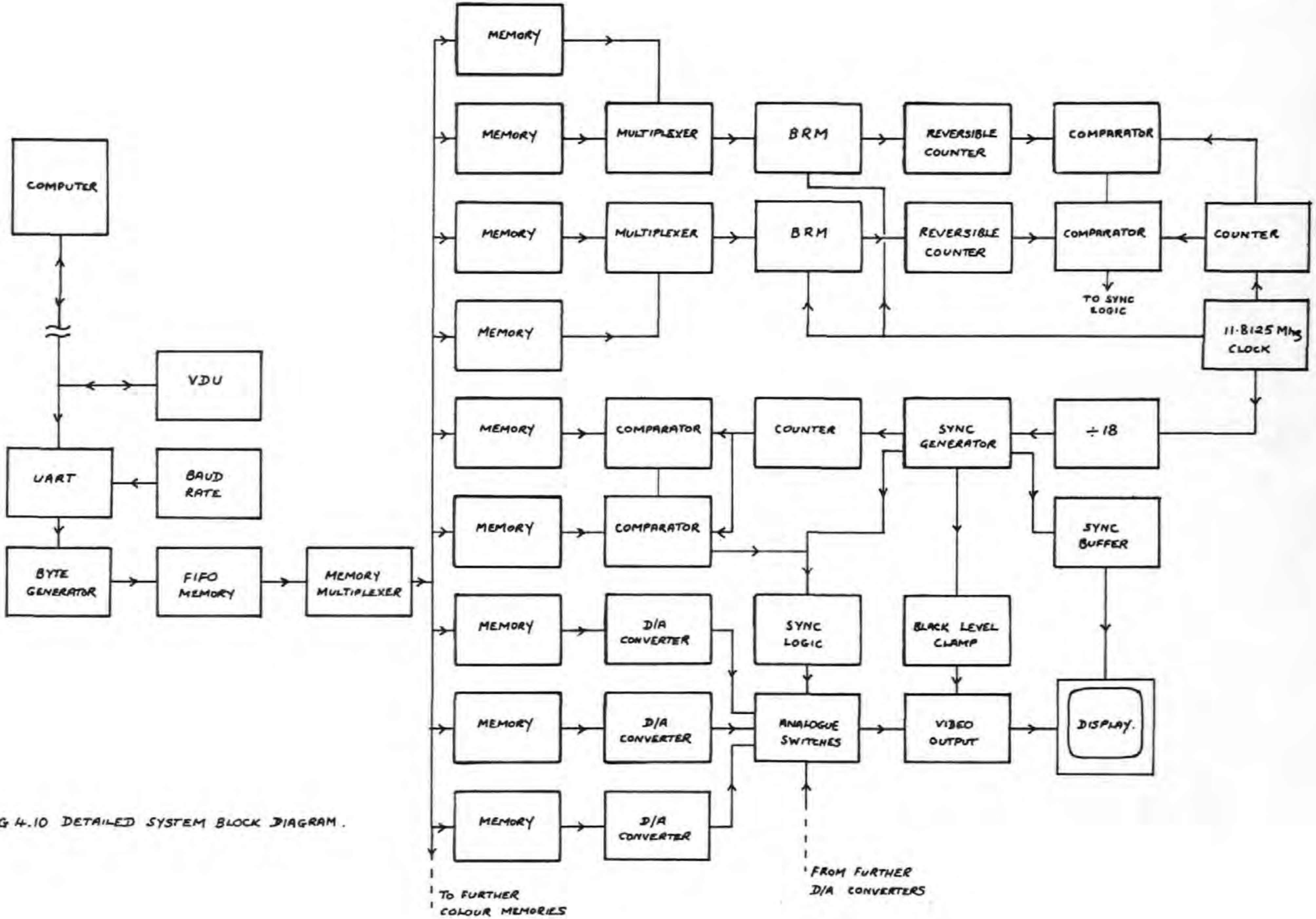


FIG 4.10 DETAILED SYSTEM BLOCK DIAGRAM.

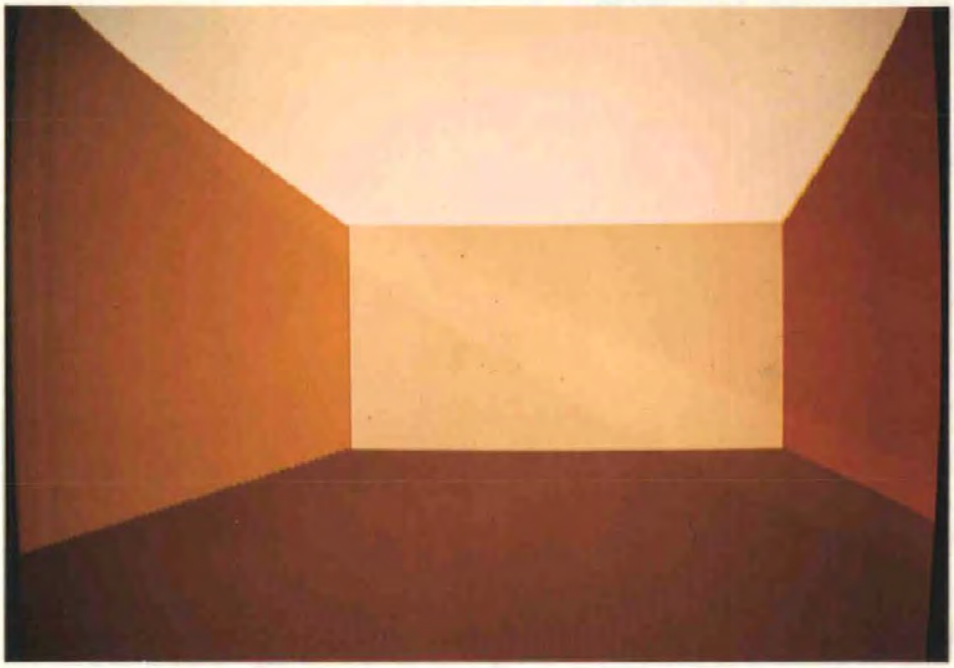


Fig. 4.11 A typical display image.

with comparator C giving an output slightly later and comparator D slightly earlier than on line 1. This sequence repeats for each scan line, producing converging vectors which define the side wall/ ceiling interfaces.

When counter B, defining vertical position by incrementing once at the end of each line, reaches the value set in memory 1, a pulse is produced by comparator 1 which inhibits counters C and D. The output from comparator 1 produces the horizontal interface between surfaces 1 and 5 and the two vertical interfaces 1-3 and 1-4. When counter B reaches the value set in memory 2 a pulse from comparator 2 reactivates counters C and D but with counter C now decrementing and D incrementing. In this way interfaces 1-6, 6-3 and 6-4 are produced. At the end of field 0, a second field blanking interval begins and the complete sequence is repeated for field 1.

4.31 Changing the viewpoint.

The gradients of the surface interfaces 3-5, 4-5, 3-6 and 4-6 are, as described in section 4.3, fixed by the rate at which counters C and D are incremented or decremented during the line blanking interval. The gradients are varied by Binary Rate Multipliers (BRMs)[2] which may be programmed to give a wide range of output count rates.

Each BRM has an input of six bits; using two devices in cascade the number of output pulses M resulting from 4096 input pulses, is defined by

$$M = L.2^{11} + K.2^{10} + J.2^{9} + I.2^{8} + H.2^{7} + G.2^{6} + F.2^{5} + E.2^{4} + D.2^{3} + C.2^{3} + B.2^{1} + A.2^{0}$$

The range of values of Ø in Fig. 4.07 is determined by the BRM input clock frequency and the count length of the BRM. The smallest non-zero value of ϕ in this case results from M = 1. With a clock rate of 11.8125 MHz, 141 input pulses are received during each line blanking interval. For M = 1, 4096 input pulses result in 1 output pulse. The corresponding value of ϕ is calculated using the following procedure.

In the standard 625 line System I, used for broadcast television in the United Kingdom, 52μs are required for one active line scan [3]. This interval corresponds to 52 x 11.8125 = 614 clock pulses. The television picture has an aspect ratio of 4:3, hence 460 clock pulses are the horizontal spatial equivalent of 575 line widths in the vertical direction. Since the raster is 2:1 interlaced, there are 287 line blanking intervals in each complete vertical transition.

The number of BRM input pulses in each field = 287 x 141. With reference to Fig 4.07,

$$\therefore \phi = \tan^{-1}\left[\frac{287 \times 141 \times M}{4096 \times 460}\right]$$

for $2\ \delta x = M_{min} = 1, \quad \therefore \phi_{min} = 1.23°$

$$M_{max} = 4095 \qquad \therefore \phi_{max} = \tan^{-1}\left[\frac{4095}{4096} \times 87.97\right]$$

$$= 89.35°$$

If the counters C and D are clocked by a fixed number of pulses during each line blanking interval, for an 11.8125 MHz clock, only 142 different gradients are possible. The BRMs are able to interpolate between these gradient values to produce a maximum of 4096 different values. To achieve this range the BRM produces an aperiodic output waveform. The BRM count sequence is continued from line to line, but counting is only enabled during line blanking intervals.

When interpolation is necessary, the number of output pulses to counters C and D will vary from line to line. Figs 4.12 to 4.15 show close-up examples of the resultant switching patterns for a range of gradient values.

4.4 Digital to analogue conversion and storage.

The inter-reflection calculation described in chapter 3 produces three numbers which define the luminance and chromaticity for each of the room surfaces, a total of eighteen numbers in all. Numbers arrive from the computer in digital form and have to be stored and converted into the analogue form required by the television display. It is necessary, therefore, to decide whether the storage should be in analogue or digital form. Analogue storage has the advantage that only one multiplexed digital to analogue (D/A) conversion is needed, but is inherently less stable, having a limited storage time. Use of digital storage allows complete control of the accuracy required with, in principle, unlimited storage time.

Given these advantages, a digital storage system using TTL latches arranged in units of eight bits is used as shown in Fig 4.16. These devices transfer information from their input to output only when an enable pulse is applied to a third terminal. Data remains unchanged at the output until a further enable pulse is received.

Having stored the picture data it is now possible to generate a video signal by a process of multiplexing and D/A conversion. As with storage, both digital and analogue methods can be used for multiplexing. Fig 4.17 illustrates the switching response time of a typical D/A converter for the 'worst case' situation in which all eight bits are changed in the same direction simultaneously. The

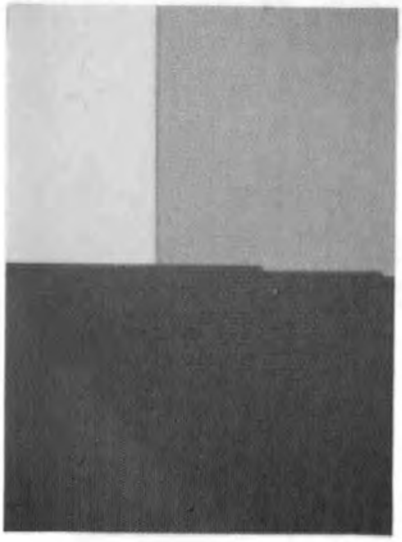


Fig. 4.12

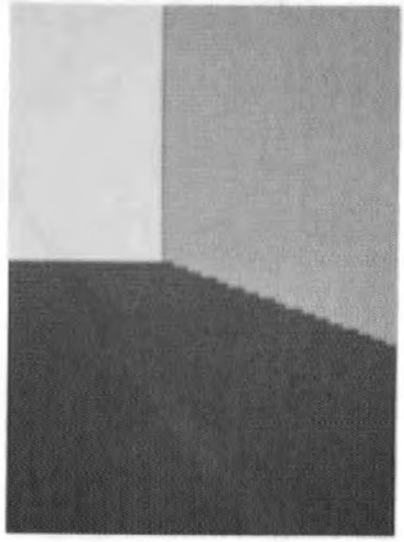


Fig. 4.13

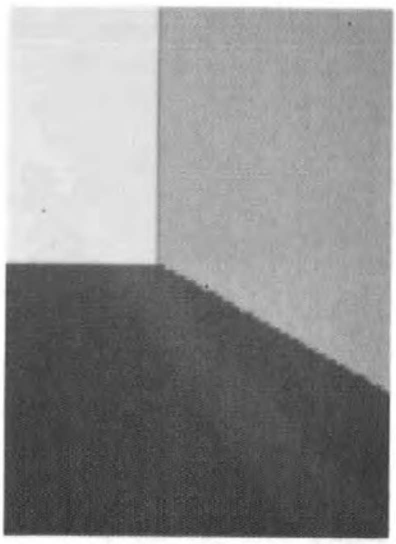


Fig. 4.14

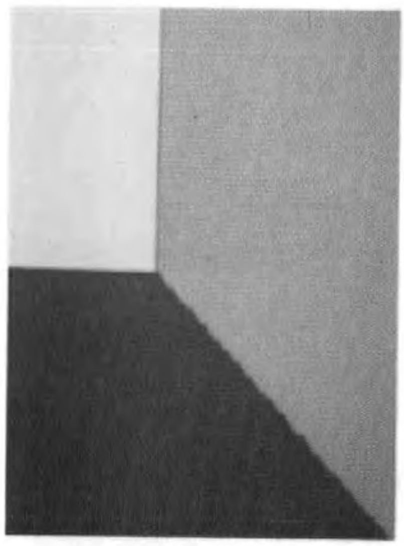


Fig. 4.15

Typical switching patterns produced by the perspective generator (magnified).

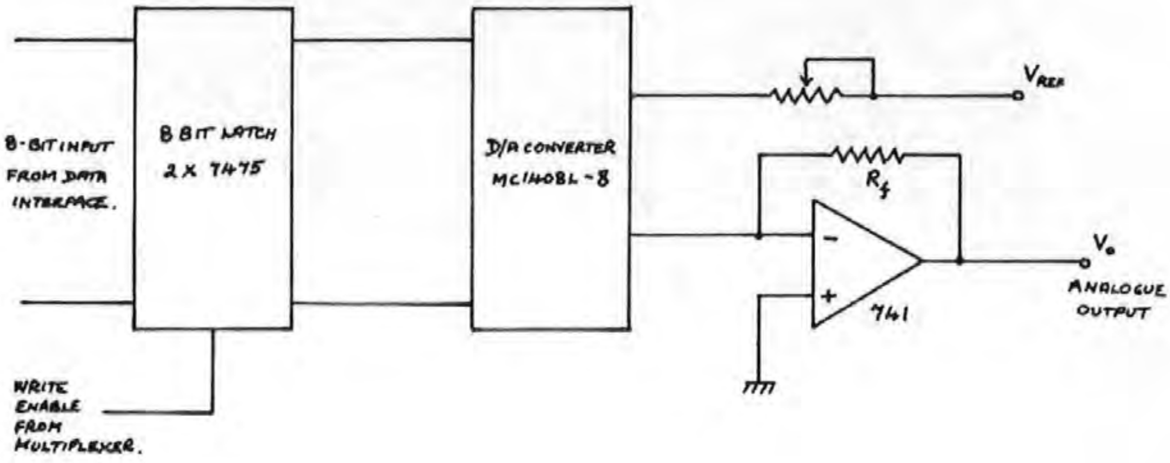


Fig 4.16 Video storage and digital/analogue conversion.

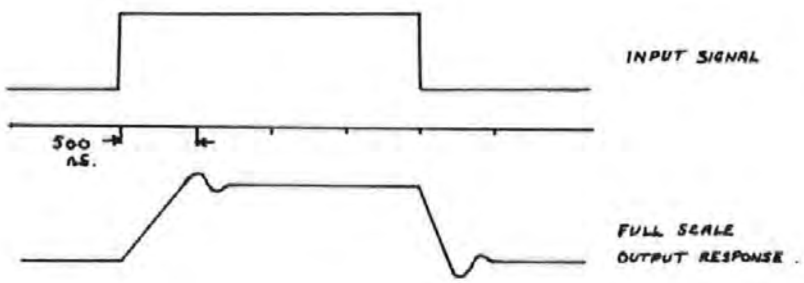


Fig. 4.17 Switching performance of MC1408L-8 converter with current/voltage conversion.

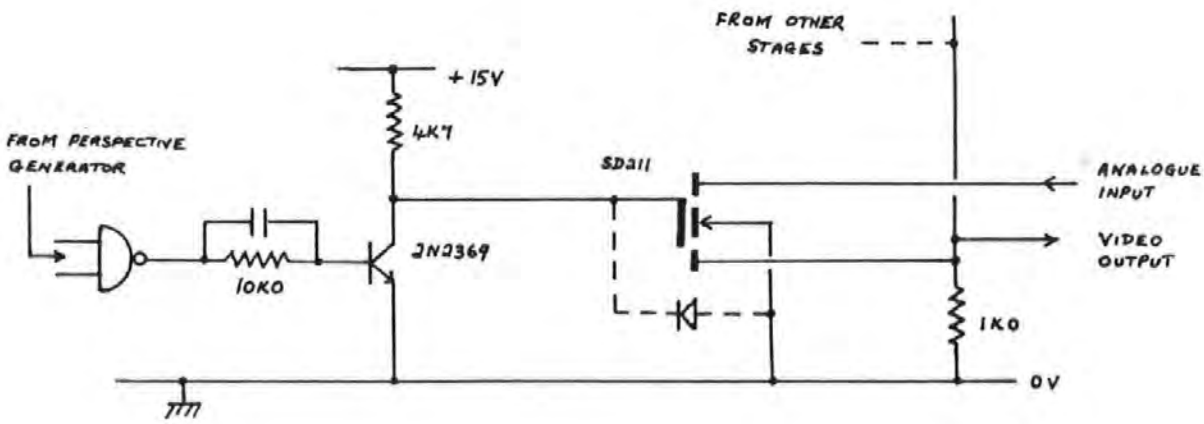


Fig 4.18 Analogue switching circuitry with TTL interfacing.

switching delay is approximately 300 ns. To produce a well defined transition on the display screen, switching times less than 100 ns are required. For this reason, if a single multiplexed D/A converter were used, a very fast device would be required along with a filtering or 'deglitching' circuit to remove the switching spikes present in the converter output. These problems may be avoided by use of dedicated slow, low cost converters for each surface in the display, switching being performed by the analogue approach as described in section 4.5.

Each latch is coupled directly to an eight-bit low cost D/A converter type MC1408L-8. This device provides a current output and is therefore buffered by an operational amplifier in a virtual earth configuration to provide voltage output. The full range output voltage is determined by the D/A reference current and the amplifier feedback resistance R_f shown in Fig 4.16. The nominal voltage maximum of 1 volt is therefore achieved by adjustment of R_f. However, because of the series resistance of the analogue switches discussed in section 4.5, each amplifier output maximum is adjusted to give 1 volt at the video output, thus giving a slightly higher level at the amplifier.

4.5 Synthesizing the video signal.

The digital comparators described in section 4.3 define position only in one dimension along either the x or y direction. Areas of the screen are defined by combinational logic using as input signals the comparator outputs as shown in Fig 4.10. This operation produces five negative-going pulses which form a time analogue of the five areas to be displayed. These pulses are used to multiplex the three analogue red, green and blue display monitor inputs so as to select the colour combination corresponding to each room surface at the appropriate time.

The red, green and blue components for each surface colour are available at the outputs of the D/A converters discussed in section 4.4. The analogue video signals are synthesized using analogue switches in an arrangement similar to the popular tri-state multiplexing used for microprocessor buses.

An ideal analogue switch in the 'on' state will give an output voltage identical in every way with the input, be capable of changing state in an infinitesimally short time and exhibit infinite impedance in the 'off'state. Practical analogue switches fall short of this ideal on every count. The devices used in this application, Signetics type SD211, do however, provide a very good approximation to this ideal. Their 'off'state resistance is extremely high ($\sim 10^{10}$ ohms) combined with a sub-nanosecond (600 ps) switching time. In the 'on' state a series resistance of approximately 40 ohms is typical. To switch the SD211 into a low resistance state requires a gate to source potential of at least 10 volts.

To interface the MOS SD211 devices with TTL gate outputs, fast NPN transistors type 2N2369 are used in the configuration of Fig 4.18 The SD211 incorporates a gate protection diode internally connected to the substrate. To avoid loading of the output by these diodes, the substrate is grounded as shown in Fig 4.18. The 10 K ohm resistors serve to limit the transistor base current to the minimum required to achieve saturation. Alone, these will cause a slowing of pulse edges due to the 'integrating' effect of stray capacitance and the transistor base-emitter capacitance. For this reason 'speed-up' capacitors are connected in parallel with the 10 K ohm resistors. Five analogue switches connected in this way are sufficient to produce the video signal for a monochrome display or for one channel of a colour display. For a complete colour system, three identical circuits are needed, providing the information for the red, green and blue channels of the display.

The series resistance of the analogue gate in the 'on' state gives rise to loss of signal amplitude, but since this is a consistent loss it may easily be compensated for in other parts of the circuit as discussed in section 4.4. This series resistance is, in fact, used to advantage. The output from the analogue gates is coupled via a 75 ohm cable to the colour display. Strictly, this cable should be terminated by a 75 ohm resistive load such that high frequency signals propagating along the line are not distorted or reflected at the termination. However, a 75 ohm load will cause significant heating of the SD211, resulting in 'on' resistance instability and possible colour inaccuracy.

This problem may be overcome by use of an additional impedance matching stage such as an FET source follower. Alternatively, the line may be operated with an open-circuited termination but driven by a source impedance of 75 ohms. Thus reflected signals are absorbed at the source. The natural 'on' resistance of the analogue switches, combined with the source impedance of the analogue voltage source described in section 4.4, are adjusted to form a close approximation to 75 ohms. Thus the analogue switches are connected directly to open-circuited 75 ohm cables, avoiding the need for additional impedance matching components.

4.6 Black level clamping

The video signal produced at the output of the analogue switches is required to change in time between five fixed d.c voltage levels. However, the colour display monitor has a video amplifier which is a.c. coupled. For this reason the incoming signal must contain information relating its amplitude to some reference level corresponding to black on the display screen.

To provide this reference level the video multiplexer is designed to switch off all the analogue gates during the line blanking interval. Thus, during this time the video signal level falls to zero, returning to normal video levels at the start of the line scan. The display monitor circuitry incorporates a clamping circuit which automatically returns the CRT signal input to a fixed 'black' level before the start of each line. Hence the d.c. level of the signal at the display tube is maintained at the value dictated by the incoming signal in spite of being processed by an a.c. coupled amplifier.

4.7 Computer interfacing.

High speed data exchange between a computer and peripheral is usually achieved by parallel data transfer using direct memory access (DMA) or programmed transfer via the CPU.[4] Experiments using a PDP8 computer showed this latter method to be satisfactory from a speed viewpoint, data for a complete picture being loaded into memory at a speed sufficiently high for the transfer to appear instantaneous to the viewer.

The disadvantage of this system is its lack of versatility. Although similar in principle, parallel data transfers from different computer types have differences in detail which require dedicated interface circuit for each type of computer. In contrast, virtually all types of digital computer provide facilities for interfacing with a teletypewriter. A display system designed to look from the computer's viewpoint like a teletypewriter thus has far wider application. The price paid for this versatility is one of speed. A standard serial data rate of 110 baud (1baud = 1 bit/sec) used by mechanical teletypewriters would require about four seconds to transfer one picture. During these four seconds a confusing flickering of displayed colour would occur as the memories are updated. This problem is partially overcome by an additional buffer memory which stores all the new data at the relatively low serial rate. At the end of this transfer, the buffer is emptied into the display store at a far higher rate (approximately 1 MHz), thus restoring a subjectively instantaneous picture change.

Computers and microprocessors are increasingly used with Visual Display Units (VDUs) which operate at far higher serial data rates (up to 9.6 K baud). Using the display system at these high rates, the time delay for serial data transfer (i.e. loading of the buffer memory) becomes negligible and performance is indistinguishable from that achieved by DMA.

4.71 Parallel interfacing to the PDP8 computer.

For high speed transfers data is taken from the computer 'Omni-bus' [5] as parallel binary words. In the case of the PDP8 these words have a total of twelve bits, eight of which are used for picture information. The required data is set up on the bus under software control using the FDIS output function available in the high level

FOCAL language.[6] In response to this command, the computer hardware produces a series of timing pulses which constitute one transfer cycle. These pulses are used to control external circuitry as illustrated in Figs 4.19, 4.20 and 4.21.

At the start of a transfer cycle the computer outputs a six bit parallel code which is used to select the particular output device. This device selection is necessary since the bus may be connected to several peripherals simultaneously. Having selected the output device, the computer transmits a sequence of timing pulses, IOP1, 2 and 3 as illustrated by Fig 4.22. The circuitry of Fig 4.21 is used to ensure that only when the appropriate data is stable on the bus can picture information be written into the display memories.

As mentioned in the introduction, the parallel data system uses hardware which is restricted to one type of computer and although offering high speed, gives further problems when transmitting data over long distances due to the high cost of the cable. Each data line must be terminated by its characteristic impedance. This low impedance places further current demands on the driving circuits increasing the cost and complexity. Failure to match impedances causes signals to be reflected from the load, a potential source of transmission errors.

4.72 Interfacing via a serial data channel

The problems associated with DMA are in part avoided when data is carried via a teletypewriter channel. Relatively low frequencies make impedance matching less critical and, since only one cable is required, long distances can be covered at far lower cost.

Data transmitted to a VDU is commonly encoded using the

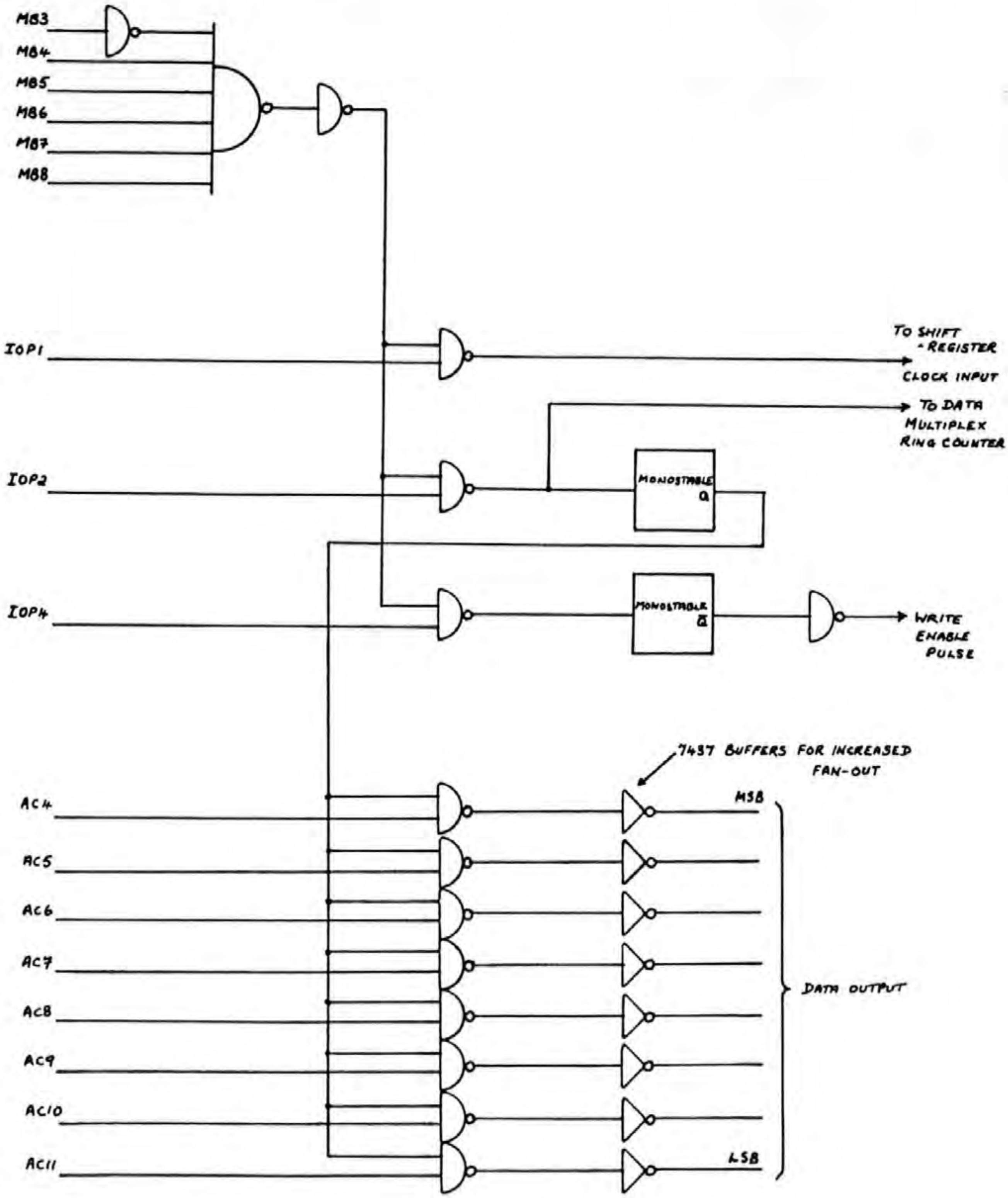


Fig. 4.19 Interface control system for PDP8 computer.

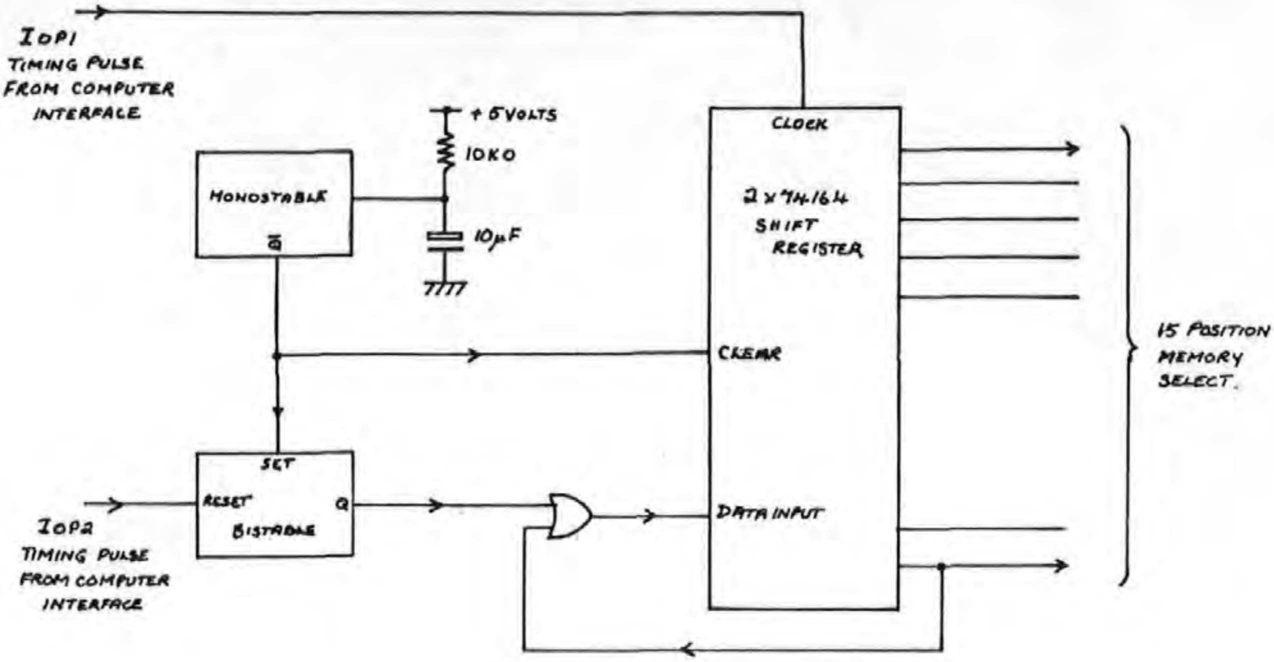


Fig. 4.20 Ring counter arrangement for memory multiplexing.

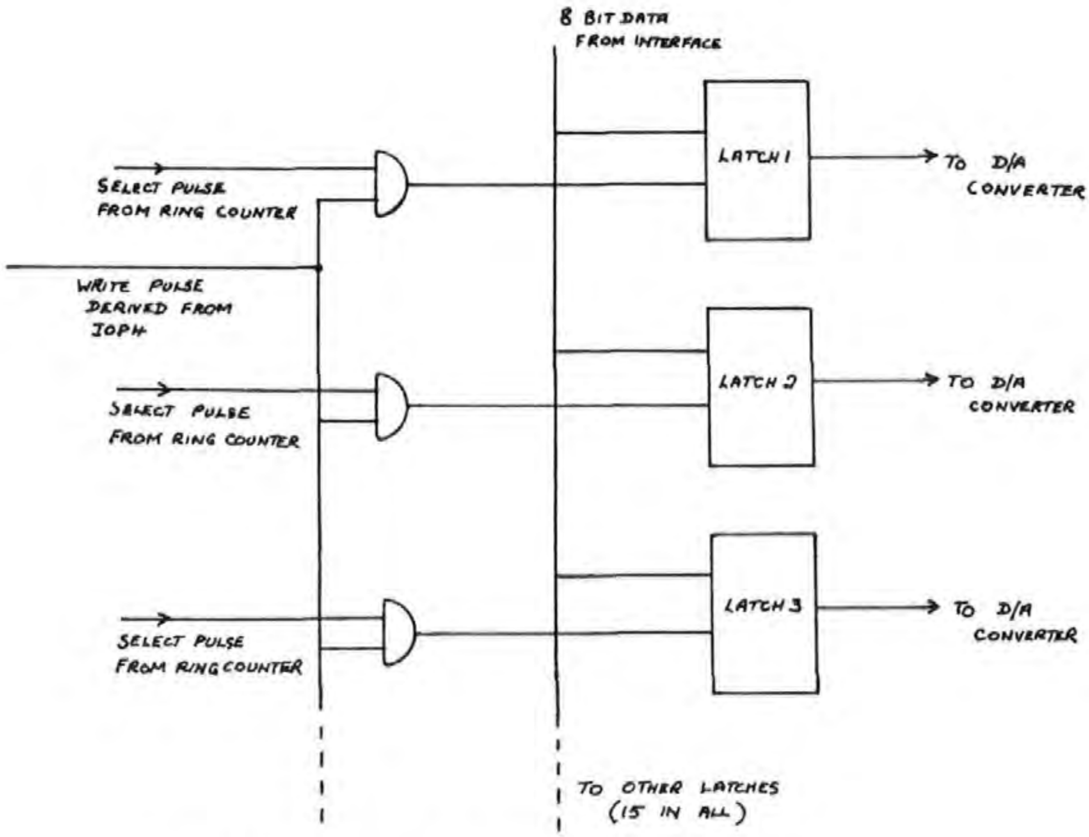


Fig. 4.21 Memory write multiplexing circuit.

American Standards Code for Information Interchange (ASCII).[7] Each character is represented by a maximum of eleven bits. Three of these bits define the beginning and the end of a character string, one bit is used for parity checking, leaving seven bits to define the character as shown in Fig 4.23. These character codes are used by the display system to represent picture information. Since the display data is in eight bit words, each one of these words is assembled from two ASCII characters sent consecutively. In the first character, the five least significant bits (LSBs) are transmitted, the following character providing the remaining three bits. The most significant bit (bit 7) is used to define which of the two characters is being sent, a logic '1' corresponding to the five LSBs and logic '0' the remaining three bits. In Fig 4.24 the complete ASCII character set is listed, each character with its equivalent decimal code.

The serial data is transmitted in the format shown in Fig 4.23 Data is converted from serial to the parallel form required by the display system by a Universal Asynchronous Receiver/Transmitter (UART) as shown in Fig 4.25. The required baud rate is fixed by a crystal controlled clock generator (type COM 5046) which allows various standard baud rates to be selected by switches. A variable RC oscillator is also available should intermediate baud rates be desired.

The serial data input may be from one of two sources; either on ASR23, 20 mA current loop interface or a V24 ±12 volt (RS232) voltage interface[8], both of which are currently popular. In the case of the current loop, the signal is converted to the required TTL

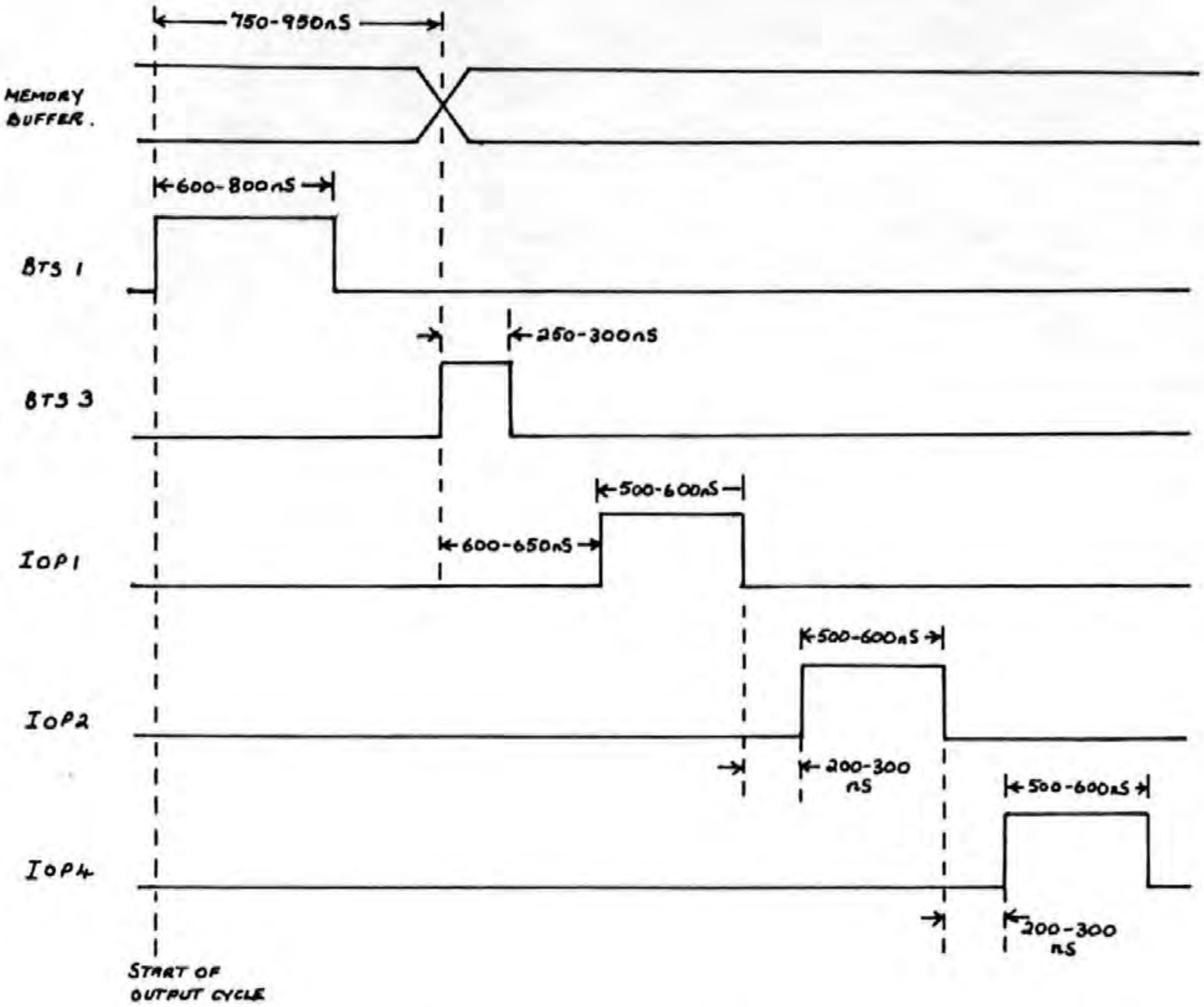


Fig. 4.22 Timing signals available to device interfaces from the PDP8 output bus.

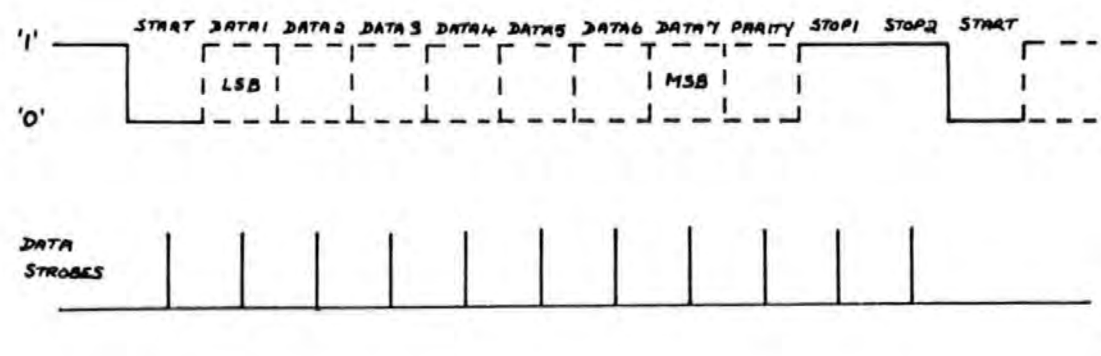


Fig. 4.23 ASCII character format.

BITS B_4	B_3	B_2	B_1	B_7 Ø B_6 Ø B_5 Ø	B_7 Ø B_6 Ø B_5 1	B_7 Ø B_6 1 B_5 Ø	B_7 Ø B_6 1 B_5 1	B_7 1 B_6 Ø B_5 Ø	B_7 1 B_6 Ø B_5 1	B_7 1 B_6 1 B_5 Ø	B_7 1 B_6 1 B_5 1
				CONTROL		HIGH X & Y GRAPHIC INPUT		LOW X		LOW Y	
Ø	Ø	Ø	Ø	NUL Ø	DLE 16	SP 32	Ø 48	@ 64	P 8Ø	` 96	p 112
Ø	Ø	Ø	1	SOH 1	DC1 17	! 33	1 49	A 65	Q 81	a 97	q 113
Ø	Ø	1	Ø	STX 2	DC2 18	" 34	2 5Ø	B 66	R 82	b 98	r 114
Ø	Ø	1	1	ETX 3	DC3 19	# 35	3 51	C 67	S 83	c 99	s 115
Ø	1	Ø	Ø	EOT 4	DC4 2Ø	$ 36	4 52	D 68	T 84	d 1ØØ	t 116
Ø	1	Ø	1	ENQ 5	NAK 21	% 37	5 53	E 69	U 85	e 1Ø1	u 117
Ø	1	1	Ø	ACK 6	SYN 22	& 38	6 54	F 7Ø	V 86	f 1Ø2	v 118
Ø	1	1	1	BEL 7 BELL	ETB 23	' 39	7 55	G 71	W 87	g 1Ø3	w 119
1	Ø	Ø	Ø	BS 8 BACK SPACE	CAN 24	(4Ø	8 56	H 72	X 88	h 1Ø4	x 12Ø
1	Ø	Ø	1	HT 9	EM 25	) 41	9 57	I 73	Y 89	i 1Ø5	y 121
1	Ø	1	Ø	LF 1Ø LINE FEED	SUB 26	* 42	: 58	J 74	Z 9Ø	j 1Ø6	z 122
1	Ø	1	1	VT 11	ESC 27	+ 43	; 59	K 75	[91	k 1Ø7	{ 123
1	1	Ø	Ø	FF 12	FS 28	, 44	< 6Ø	L 76	\ 92	l 1Ø8	| 124
1	1	Ø	1	CR 13 RETURN	GS 29	– 45	= 61	M 77	] 93	m 1Ø9	} 125
1	1	1	Ø	SO 14	RS 3Ø	. 46	> 62	N 78	Λ 94	n 11Ø	~ 126
1	1	1	1	SI 15	US 31	/ 47	? 63	O 79	_ 95	o 111	RUBOUT (DEL) 127

Fig. 4.24 Complete ASCII character set.

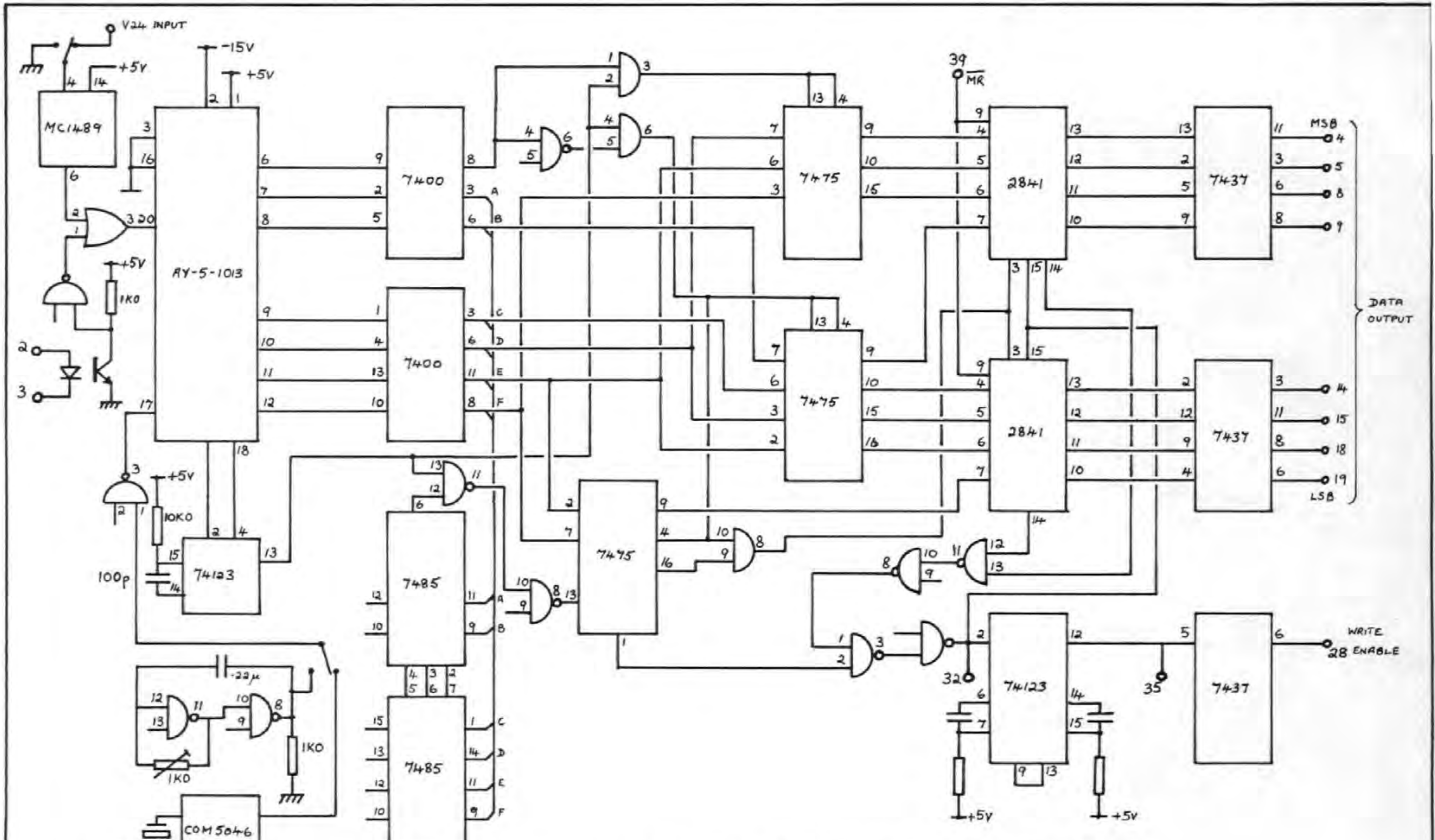


Fig. 4.25 Serial to parallel data interface.

voltage levels by an opto-isolator. In the case of the V24 signal, the ±12 volts input waveform is translated to TTL levels by an MC1489 buffer.

The ASCII character set shown in Fig 4.24 contains codes which are available for purely control functions. Two of these codes have particular significance for the display system. The data input is designed to operate in parallel with a teletypewriter or VDU, the UART has no means of discriminating between data for the VDU and that containing display information. For this reason the serial display interface is designed to operate in two modes. When control character 31_{10} is output by the computer, this is recognised by the interface circuitry and a latch is set switching the system into the standby, non display mode. Subsequent characters received by the interface are ignored, to be used only by the VDU. It is usual for computer programmes used by the display to output this control character as part of an initial set up procedure. A second control character 29_{10} performs two functions, switching the display interface into the display mode and disabling the VDU. Subsequent ASCII characters transmitted by the computer are interpreted as display data.

In the circuit of Fig 4.25, a digital comparator is used to detect the presence of either control character. Since these differ only by one bit, this bit is latched to act as a mode switch. Bit seven is used to multiplex the UART output to one of two latches, allowing an eight bit word to be assembled from two consecutive characters. Since the rate of word generation will vary directly with the baud rate, a buffer memory is used to ensure that a change of displayed information will always appear subjectively instantaneous.

This buffer is a first-in, first-out (FIFO) memory which allows input and output functions to be performed asynchronously.[9] As each eight bit word is assembled, a pulse is sent to the FIFO to shift the data into the memory. Internal circuitry now automatically causes the word to 'ripple through' the memory to the final location. Subsequent words are assembled in consecutive memory locations. When a complete set of data for one picture has been stored, control character 31_{10} is sent, switching the interface system out of the 'display' mode and reactivating the VDU. This control signal also activates the 'shift out' circuitry of the FIFO, allowing data to be transferred at high speed ($\sim$1 MHz) to the display memory, producing an apparently instantaneous picture change.

The interface can be used with transmission rates ranging from 110 baud, giving a maximum delay of four seconds during transfer, to 9.6 K baud which gives negligible delay.

4.73 Interface output multiplexing.

For both parallel and serial interfaces it is necessary to multiplex the data output lines such that each word arrives at the appropriate display store.

The circuitry illustrated by Fig 4.20 and 4.21 is common to both serial and parallel cases, differing only in the source of control signals. Power switch-on produces a delayed output pulse from a monostable multivibrator which is used to clear a shift register and preset a bistable which then sets a logical '1' at the shift register data input. For the parallel case on FDIS instruction causes a sequence of pulses to be output from the interface as illustrated by Fig 4.22 Timing pulse IOP1 is used to clock the shift register whilst IOP2 serves to reset the bistable. Thus, after switch-on the first FDIS execution causes a '1' to be shifted into the first output position of the shift register and the data input to be set to '0'. Subsequent FDIS instructions cause the '1' output to be cycled round the shift register, steering the data to fifteen latches selected in sequence.

To ensure reliable data storage it is necessary to provide a short write pulse occurring only whilst data is stable. This pulse is derived from IOP4 as shown in Fig 4.19. The minimum write pulse length required by 7475 latches is 100 ns. A pulse of 200 ns is used for this purpose as shown in Fig 4.19. Since a monostable feeds write pulsed into fifteen TTL inputs it is necessary to use a buffer (type 7437) to provide the necessary fan-out.

In the case of the serial interface, data output storage begins when a pulse is received signifying the switching of the interface from 'display' mode to the standby mode. At this instant the FIFO has stored a complete picture and is ready to transfer data to the display memory. This pulse clocks the shift register, as did IOP1 in the parallel case and also results in two delayed pulses from cascaded monostables which perform the same function as IOP2 and IOP4 respectively. The second and subsequent write cycles are initiated by an 'output ready' pulse from the FIFO buffer. The output sequence is repeated until the final word is read from the FIFO, after which no further 'output ready' pulses are received, signifying completion of the picture transfer.

Detailed circuit diagrams for the display system are given on pages 167 - 172 inclusive.

References

1. Lappalainen, P.
Digital Circle-Pattern Generator for T.V.
I.E.E.E. Transactions on Broadcasting. vol. BC-19 no. 3 (1973).

2. The TTL Data Book for Design Engineers.
Second edition. Texas Instruments Ltd., p. 7 (1976)

3. Characteristics of Monochrome T.V. Systems.
Report no. 308-2, CCIR Conclusions, 5, part 2, (1970).

4. Zissos, D. and Duncan, F.G.
Digital Interface Design, United Kingdom Atomic Energy Authority Research Group, Harwell Series. Oxford University Press (1973).

5. PDP8/L Small Computer Handbook. Digital Equipment Corporation.

6. Programming Languages. PDP8 Handbook Series.
Digital Equipment Corporation.

7. Nashelsky, L.
Introduction to Digital Computer Technology. Wiley p. 93, (1977).

8. EIA Standard RS-232-C and Application Notes.
Electronic Industries Association, Washington D.C. USA, (1969).

9. MOS Memory Data Book,
Fairchild Camera and Instrument Corporation. (1976).

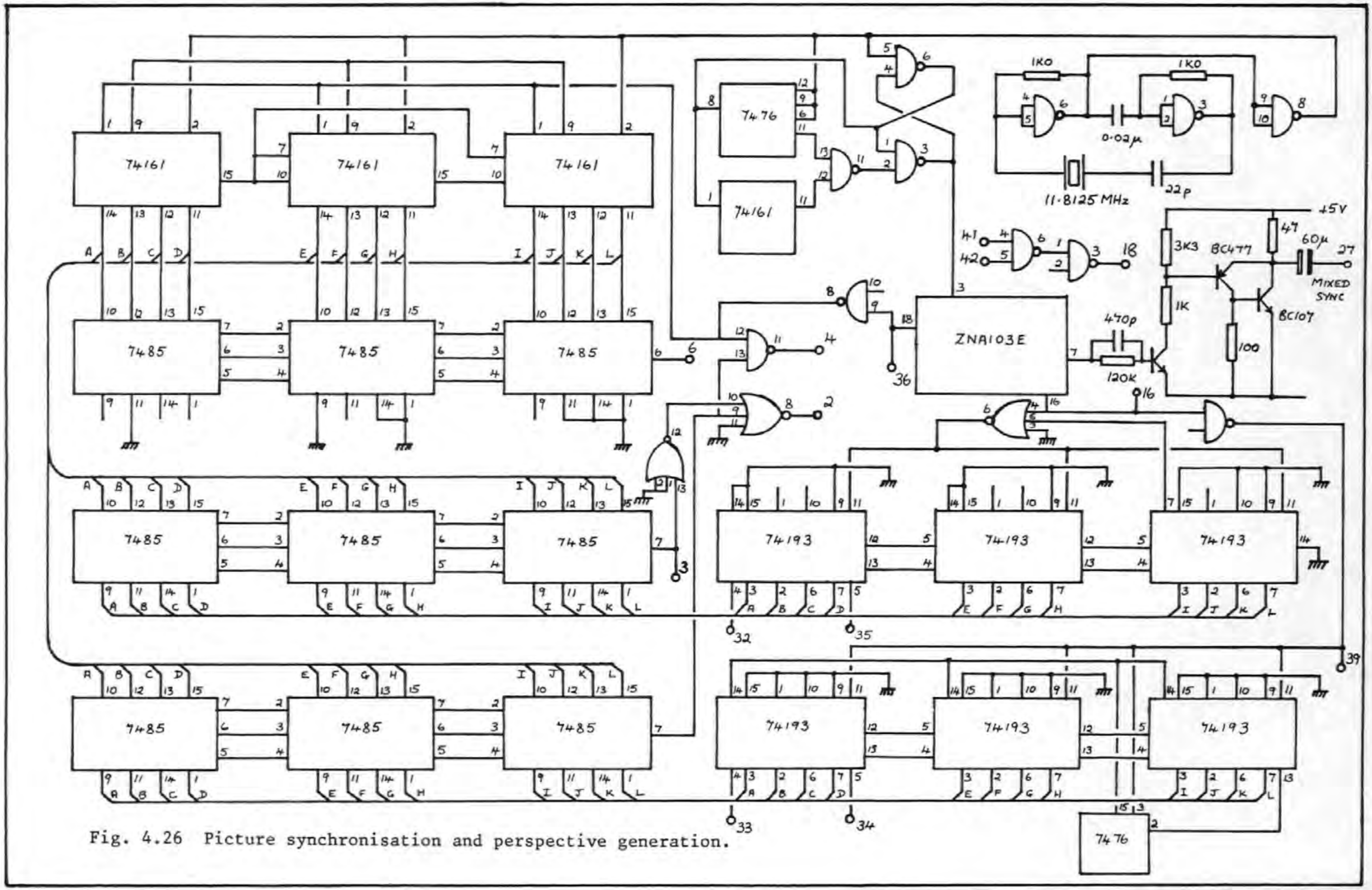


Fig. 4.26 Picture synchronisation and perspective generation.

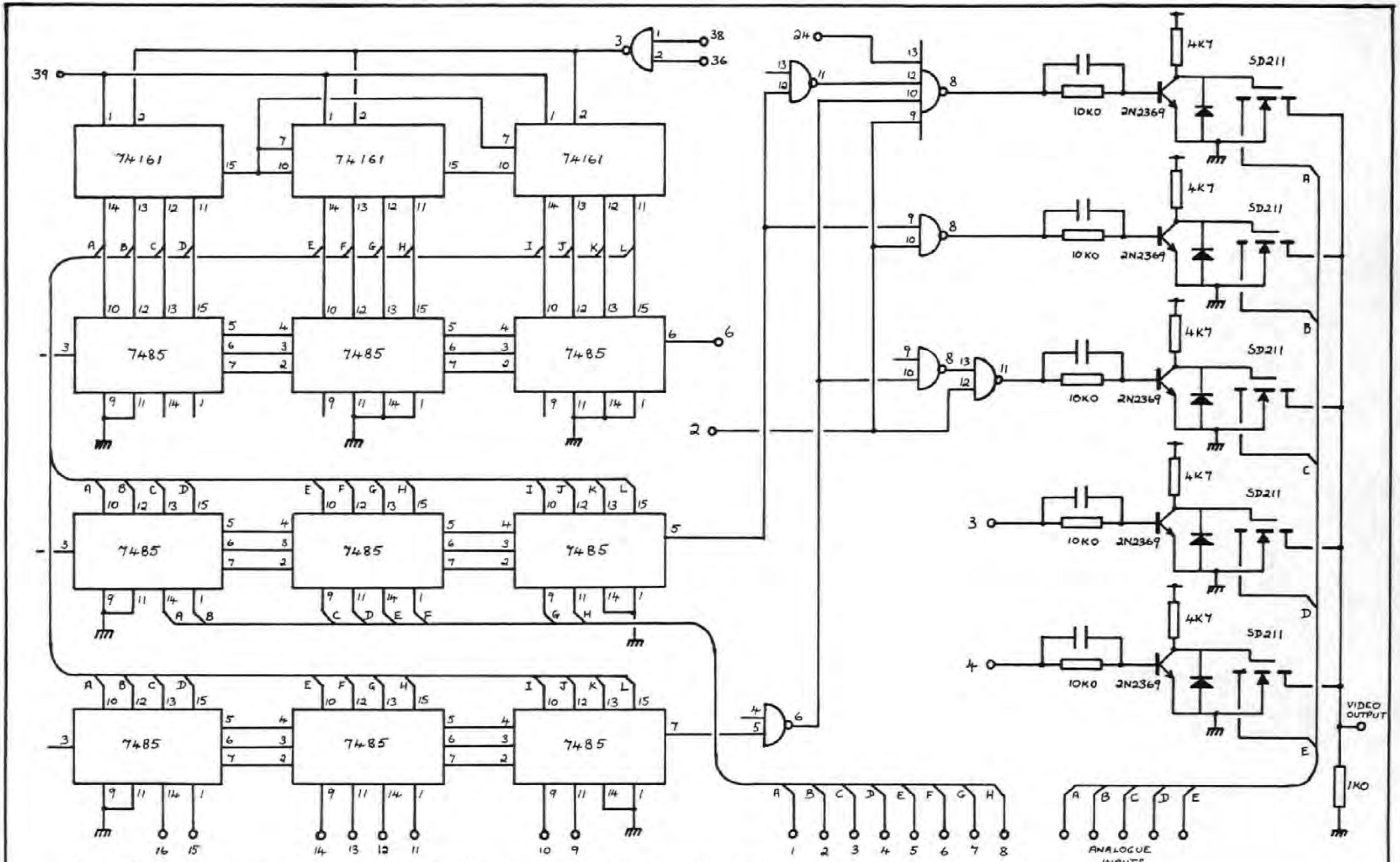


Fig. 4.27 Picture vertical positioning and analogue multiplexing.

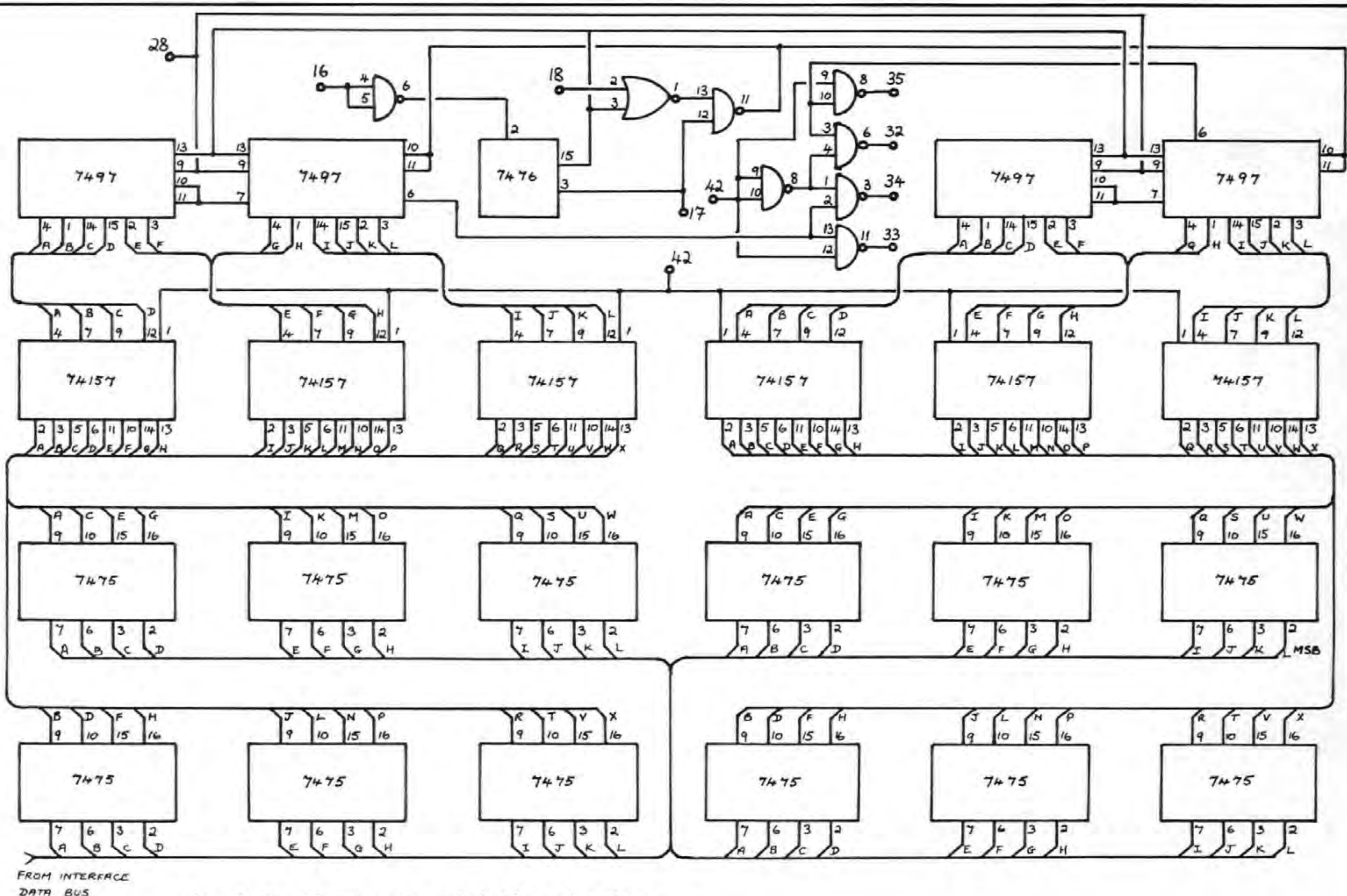


Fig. 4.28 Binary rate multiplier circuitry.

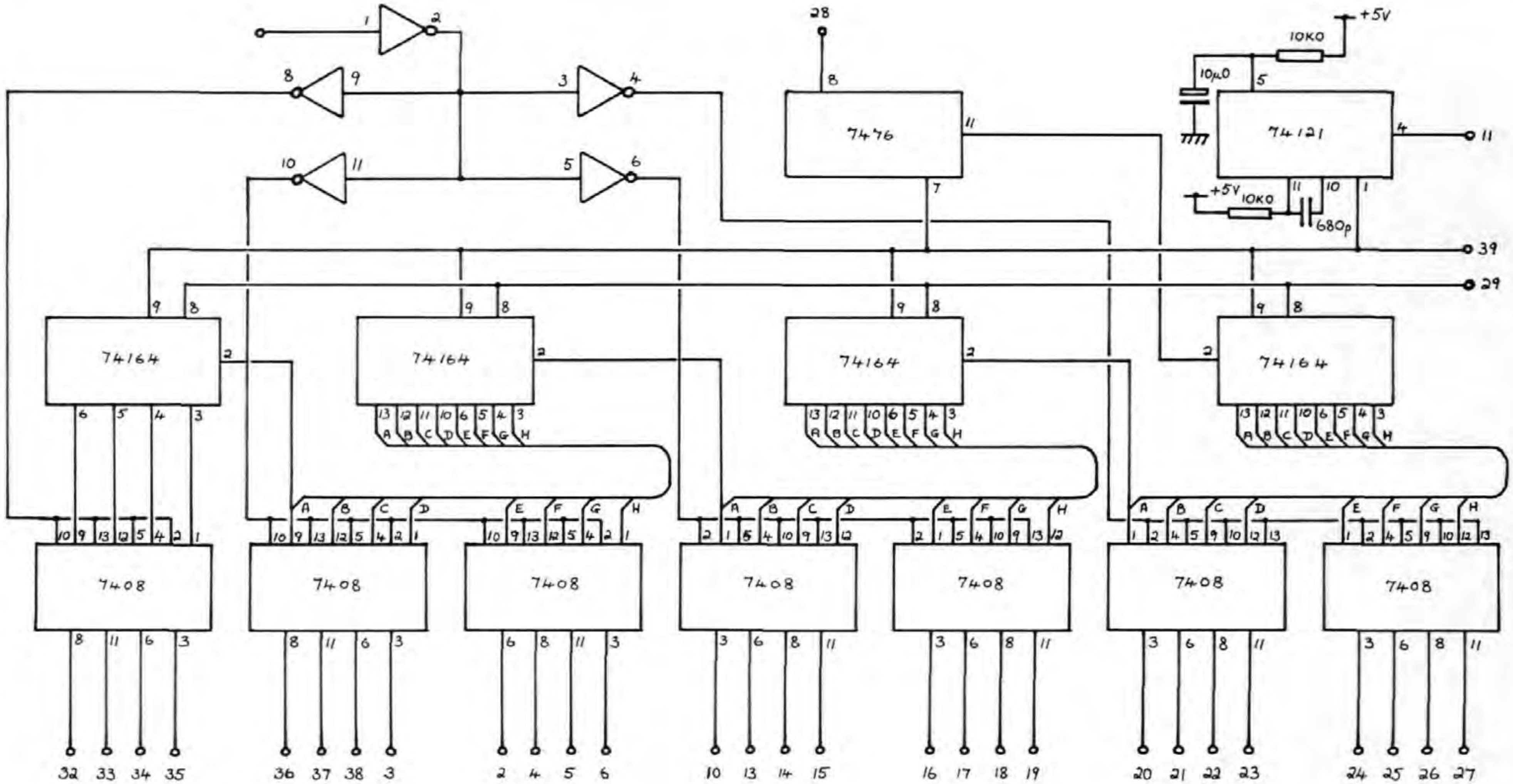


Fig. 4.29 Memory write multiplexer.

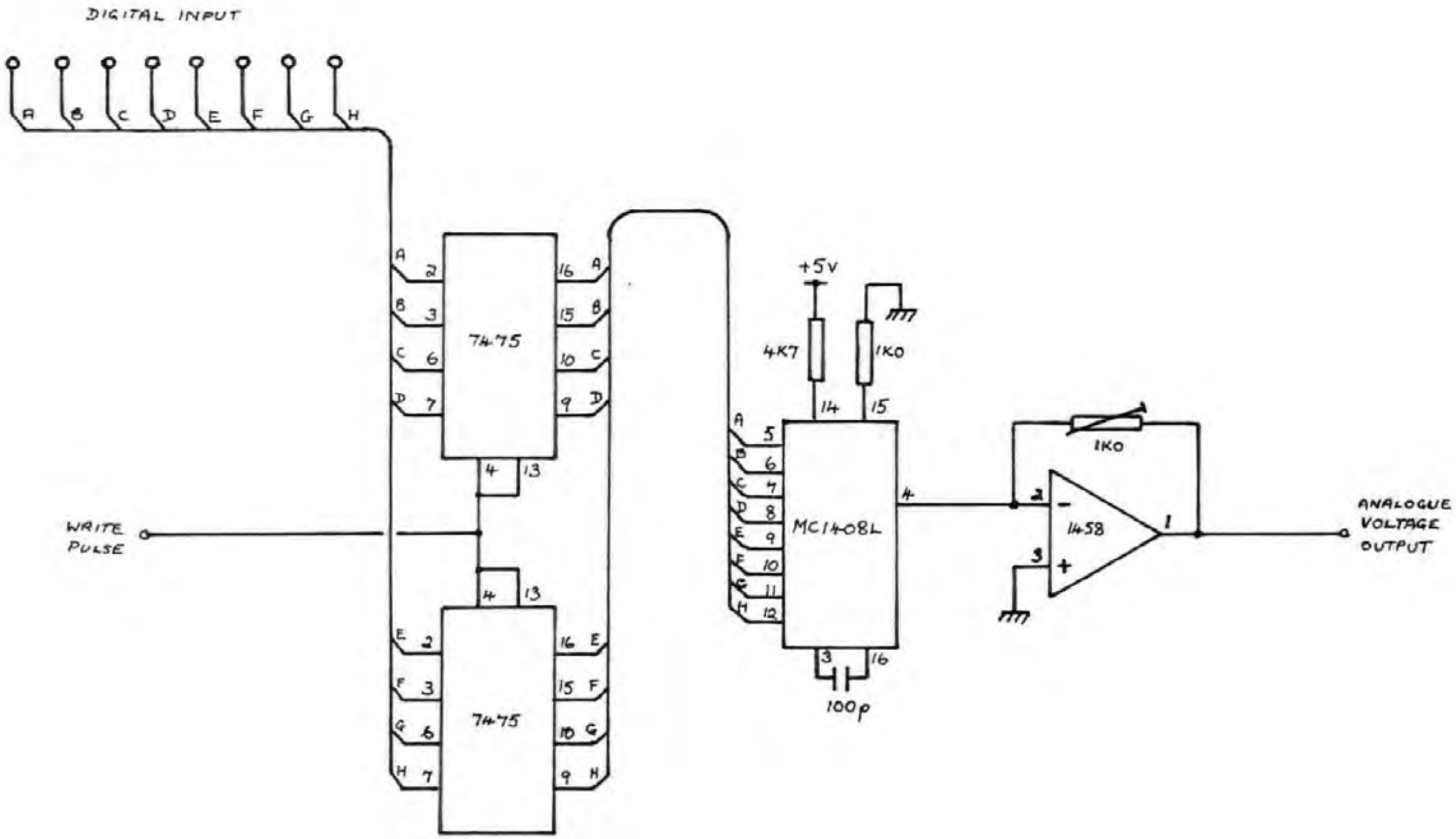


Fig. 4.30 D/A converter arrangement.

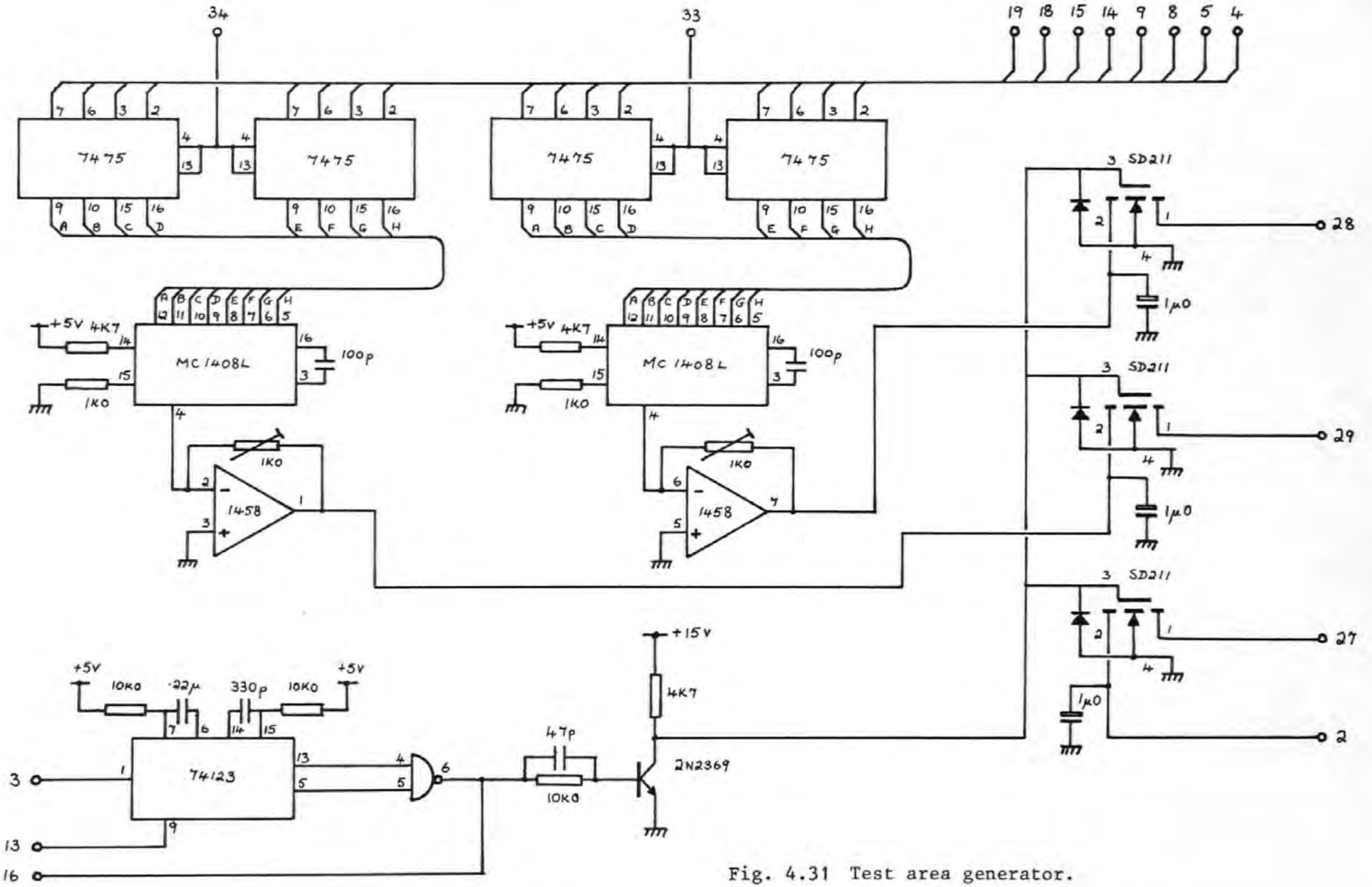


Fig. 4.31 Test area generator.

Chapter 5

COLOUR PERCEPTION AND COLOUR DISPLAYS

The colour display system described in chapter 3 was designed to reproduce the chromaticities and relative luminances of a real interior. This approach was taken as a practical starting point rather than as an absolute display method since it is well known that photometry and colorimetry alone do not provide a reliable measure of colour sensation.

Chromaticity reproduction can provide certainty of reproducing colour sensation only in the hypothetical case in which every detail of the original scene is duplicated with identical viewing conditions. Practical colour reproduction necessarily involves a change in viewing conditions and some loss of information. In the case of broadcast television, for example, exact chromaticity reproduction has not proved to be the most appropriate approach, a point developed further in section 5.3

This chapter examines the possible shortcomings of chromaticity reproduction as a display objective. Finally, the possibility of predicting perceived colour is discussed.

5.1 Colour appearance and the standard observer.

In 1943 Kelly[1] published data which attached colour names to various areas of the CIE chromaticity diagram. These names were used to describe coloured lights viewed against a dark background, experimental conditions similar to those used to derive the standard observer data on which the CIE colour system is based. For these simple viewing conditions chromaticity co-ordinates can provide some measure of colour appearance and an observer could be trained to name isolated stimuli consistently. If the viewing conditions are changed, for example, to a surface colour of the same chromaticity against a light background, the colour sensations will also change resulting in a different name being given for the same stimulus.

In a complex visual field, the matching of two separate coloured elements is no longer predicted by a simple measure of luminance and chromaticity. As will be shown in the later sections of this chapter, the measurement of colour sensation requires more sophisticated instruments than a photometer and colorimeter.

5.2 Colour appearance in a complex field.

For the general case of a complex image, colour stimuli of varying luminance are juxtaposed in space and time. Temporal changes occur even for stationary images, due to involuntary eye movements. In these conditions, the appearance of colour stimuli becomes a function of their environment. These perceptual changes are believed to result from sensitivity changes within the visual system.

In the context of brightness perception the effect of temporal contrast is described in the literature as adaptation or successive contrast. Spatial effects have been called simultaneous contrast and induction, and are also sometimes described as adaptation. In this thesis temporal and spatial contrast are referred to as adaptation and induction respectively, the corresponding colour contrast effects being chromatic adaptation and induction.

In addition to the perceptual effects of induction and adaptation, there is evidence that less tangible effects, such as memory, may influence perception in complex fields. Colour perception may well be the result of processing on at least three levels, involving the retina, the neural pathway to the brain and the brain itself. Although an exhaustive study of the extensive literature covering the theory of colour perception is outside the scope of this thesis, a review of some relevant studies is given in Appendix 5 .

5.3 Television colour reproduction and colour perception.

5.31 Display criteria for broadcast television.

Television images are subject to the effects of adaptation and induction inherent in the perception of complex scenes. It could be argued that contrast effects will be the same in the reproduction as in the original image and may therefore be ignored. However, there are important practical differences. The reproduced scene is two-dimensional and usually has a smaller angular subtense and lower luminance than the original. It is bounded by borders, beyond which the surroundings bear no relation to the original scene. These practical differences suggest that reproducing chromaticity may not necessarily be the most appropriate criterion for reproducing colour sensations via a television display.

In broadcast television it is the convention that a standard white surface is reproduced in the display with a fixed chromaticity, regardless of the colour of the original illuminant. This adjustment, known as matrixing, is similar to chromatic adaptation of the eye, which tends to maintain the appearance of white approximately constant. Although matrixing is often done manually, attempts have been made to automate the process (Taylor, 1971[2]; McMann et al, 1974[3]). In the European PAL system I, the chromaticity used for the reproduced white, known as the reference neutral point, is that of illuminant D_{65}. Only when the original scene illuminant matches D_{65} could exact reproduction of chromaticity be even a possibility. For other reasons, however, chromaticity is not reproduced exactly, even in this special case.

Exact reproduction of chromaticity requires that a linear relationship is maintained between the transmitted colour signals and those presented by the display. Practical display tubes have a non-linear relationship between output luminance (L) and input voltage (V), as shown in Figure 3.13. This relationship, of the form $L = aV^{\gamma}$ where a and γ are constants, would cause errors both of hue and saturation if left uncorrected. Since the luminance range of television displays is restricted, colours are often reproduced at a lower luminance than in the original image. Low luminance is known to be accompanied by a loss of saturation. Bartleson and Breneman (1967)[4] showed that a reduction of apparent contrast and colour saturation in an image is to be expected as the surround illumination is reduced. Since television images do have relatively low luminance and are generally viewed with low surround luminance, it seems desirable that the reproduced saturation is greater than that of the original scene. This result is achieved by under-correcting for the display non-linearity, giving an overall system gamma of 1.25. The resultant saturation increase is accompanied by some hue error (Brown, 1971[5]).

In broadcast television, therefore, exact reproduction of the colorimetric properties of the televised objects has not proved to be either necessary or even desirable, to produce the best subjective result. The viewer rarely compares a television reproduction directly with the original scene. There is evidence to indicate that the preferred reproduction chromaticities for colours known to have a strong memory reference, such as foliage and human flesh tones, can differ appreciably from their original measured values (Bartleson and Bray, 1962[6]; Bartleson, 1966[7]; Sugimoto et al, 1973[8]). Bartleson (1968)[9] states that 'It is enough that a colour television picture be believable, that it presents images that are pleasing to the viewer'.

5.32 The choice of the reference neutral point for coloured displays.

As mentioned in section 3.51, television systems use chromaticity as the criterion for the reproduction of white. Any white-appearing object colour placed before the camera should be reproduced with a chromaticity close to a reference neutral point matching illuminant D_{65}. The choice of reference D_{65}, whilst not arrived at by accident, seems to result from conflicting requirements and has been the subject of some controversy (Townsend, 1962[10]).

In the display of monochrome television, the viewers' tolerance to the 'white' point is large. Experimental evidence shows a preference for a cold or bluish white rather than warmer or sepia tones which are equally practicable (Carnt and Townsend, 1957[11]). It may be that a colour which forms a contrast with the surround illumination is more attractive.

Typically, colour temperatures of 11,000°K are used for the phosphors in monochrome CRTs. In the early days of colour television broadcasting, a large proportion of material was still broadcast in monochrome. By using a colour balance point close to the monochrome 'white', the reproduced monochrome material would appear in the familiar tones of the monochrome-only receivers.

The phosphors used in colour CRTs have, in general, low luminance output compared to monochrome phosphors. A series of psychophysical experiments by Hurvich and Jameson (1951[12]) revealed that 'white' or neutral perception is a function not only of the size and chromaticity of the stimulus, but also its luminance. These results, illustrated in Figure 5.01, showed that given time for chromatic adaptation of the viewer, a wide range of colour temperatures could evoke the sensation of white. For a particular colour temperature there is a threshold luminance below which white is no longer perceived. For adaptation to stimuli in the range 2500°K to 10000°K a medium colour temperature (around 6500°K) is the most efficient in terms of its 'white producing' capacity. On average, a stimulus at 6500°K would require lower luminance to produce perception of white for observers adapted to colours within the above range.

For colour television displays in which the available luminance is restricted, a normalised white colour temperature close to 6500°K would seem to offer obvious advantages, at least from an engineering viewpoint. In America, a chromaticity matching illuminant C (6740°K) was selected in 1953 as the reference white for the NTSC television system. According to Townsend (1963) this choice by an NTSC panel was based on personal

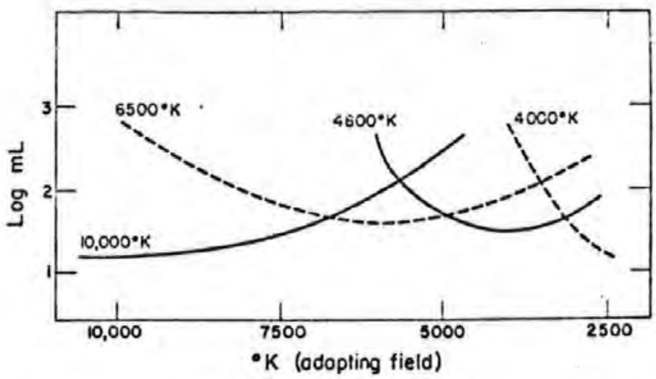


Fig. 5.01 Luminance thresholds for a white sensation as a function of colour temperature for various adapting illuminant colours (1 second stimulus with an 11.7° target.

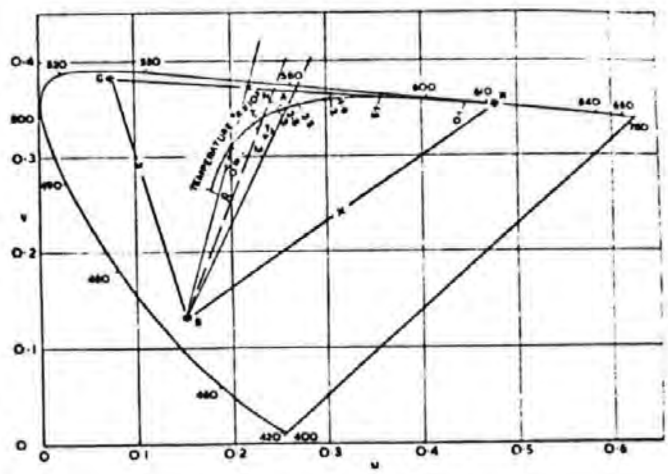


Fig. 5.02 Construction to find the best approximation to reference white points, matching illuminants A and C by adjusting only the level of the blue display primary.

opinions rather than objective arguments. This decision could also have been influenced by colour preferences for interior lighting which are known to vary with climate and latitude (Winch, 1949[13]).

MacAdam (1955)[14] reviewed the problem of television colour balance, paying particular attention to chromatic adaptation produced by the ambient lighting normally present in the viewing environment. From experience in the display of colour films and transparencies, MacAdam recommended that ideally two neutral points should be used for television, one close to 2700°K for evening viewing in tungsten lighting and the other around 6700°K to be used during daylight hours. Should two neutral points prove technically too complex, he suggested a compromise point of 4000°K, giving minimum perceived neutral shift with changes in ambient conditions.

A further study by Townsend (1962)[10] used 100 observers to determine the preferred screen colour temperature for various ambient lighting conditions. The surround illumination was shown to affect the viewers' perception both by changing chromatic adaptation and by changing the measured chromaticity of the displayed stimuli.

With the normalised picture white set to match illuminant C and surround lighting also of this chromaticity, the ambient illumination produces a change only in the purity of the colours. In general, this type of chromaticity shift is subjectively acceptable (Pearson, 1975[15]) and may be simply offset by an increase in chrominance gain. If a surround illuminant not matching the picture neutral is used, for example incandescent lighting with a standard 'NTSC' or 'PAL' display, changes not only in purity but also in hue may occur. This type of chromaticity error is less acceptable, particularly when displaying human flesh tones.

Neglecting the effect of chromatic adaptation, Townsend argues a strong case for matching the neutral point of the display to that of the surround illuminant. The use of a 'floating' neutral point offers an ideal solution to this problem, but recent broadcasting experience apparently has not shown this additional system complexity to be justified. Like MacAdam, Townsend suggested an engineering compromise in which only two neutral chromaticities are used, one close to that for illuminant A and another approximating to illuminant C. He suggested a further simplification in which the neutral point is changed by switching the amount of one display primary between two levels. This simply moves the neutral point along a straight line passing through the primary chromaticity. As can be seen from Figure 5.02, the relative chromaticity positions of illuminant A and C show that either the red or blue primary could be switched for this purpose. Generally the red phosphor has a relatively low efficiency and it is desirable to keep its output level as high as possible. For this reason the blue primary level is suggested as the most appropriate for switching the display neutral point.

In order to include the effects of viewer adaptation, Townsend designed an experiment using 100 observers, to determine the preferred display 'white' for particular ambient lighting conditions. The neutral point was switched between three positions corresponding to what might be associated with a warm, neutral and a cold colour balance respectively. Three background illuminant colours were used, each at three different luminance levels. The various combinations of display and surround stimuli were presented randomly, care being taken to allow time for complete adaptation. Of the two extreme viewing conditions, one with surround illuminant C and display neutral point matching illuminant A, the other with an illuminant A surround and illuminant C neutral point, the former condition was preferred. This finding is significant since both the NTSC and PAL television standards use a normalised 'white' point approximating to that of daylight and most television viewing takes place with ambient lighting giving a surround colour close to that of illuminant A.

In order to minimise apparent changes of colour balance with changes in ambient lighting, Townsend recommends a compromise neutral point having a colour temperature of 3500°K (Figure 5.02). Subjectively, this white is closer in colour to illuminant A than to illuminant C, consistent with viewers' preference for a slightly 'warm' white for colour pictures and with the probability that viewing will be done under ambient lighting approximating to the colour of illuminant A. Hunt (1975)[16] states that a television display with a reference white close to the colour temperature of tungsten light appears intolerably yellow when viewed in ambient daylight, but that a white point matching a daylight chromaticity, for example D_{65}, does not appear intolerably bluish when viewed in ambient tungsten light. This difference, it is claimed, is explained by the difference in surround luminance levels, that with daylight tending to be far higher, on average, than from artificial illuminants.

If ambient viewing conditions are subject to change, the choice of reference white chromaticity is inevitably a compromise. There is general agreement that the effects of contrast with the surround are minimised if the display luminance is kept as high as possible. In this way the display screen can more easily dominate the viewer's state of chromatic adaptation. Higher luminance can be obtained by using display phosphors with lower purity. Broadcast television experience (Hunt, 1975)[17] shows evidence that high luminance, lower purity phosphors, whilst degrading the display performance in terms of chromaticity accuracy, produce a result which is preferred subjectively.

5.33 The significance of the achromatic point.

There is a general agreement in the literature that the perception of colour is closely related to, and may be predicted from, a knowledge of the chromaticity of a stimulus which at the particular instant appears neutral. This chromaticity is usually called the achromatic point. In colour reproduction, another chromaticity, the colour balance point, is also relevant. The colour balance point is a weighted average chromaticity of all the picture elements in the reproduction.

Pearson (1975)[18] suggested that where a display is viewed in a dark room or with low ambient illumination, the achromatic point is close to the colour balance point. If the colour balance point also corresponds to the reference neutral, colours perceived in the display should have approximately the same appearance as those in the original scene. For a display in which the colour balance point differs from the reference neutral chromaticity, such as the two primary system investigated by Pearson and Rubinstein (1971)[19], the achromatic point is likely to differ from the reference neutral chromaticity. In this case the position of the achromatic point must be known before perceived colour can be predicted.

The importance of the achromatic point for colour perception has been emphasized by Judd (1940)[20], Hunt and Winter (1975)[21], Richards and Parks (1971)[22], Pearson (1975)[23] and Wright (1981)[24]. Pearson defined the achromatic point as approximately the chromaticity of a small test area with brightness equal to the average of the visual field which, for the particular viewing conditions, appears neutral. Hunt and Winter (1975) used colour naming of a test area with luminance equal to that of the adapting field, as a measure of the achromatic point.

Judd (1940) showed that having determined the achromatic point, the colour perceived from a given stimulus could be represented in a chromaticity diagram, as shown in Figure 5.03, by a parallel translation of the vector $n_o s$, between the achromatic point and the stimulus chromaticity, to the position na where n is the reference neutral point and a the chromaticity of an aperture stimulus with the same brightness and colour appearance as s in the complex field. Point a is known as the apparent or perceived chromaticity.

The colour display described in the earlier sections of this thesis provides a graphic image in which surface colours are reproduced as colorimetric matches for those in a real interior, taking into account the colour rendering properties of the light source. The colour balance point is therefore subject to variations and may differ significantly from the reference neutral point. The findings of Pearson suggest that in this case the achromatic point may also change, particularly if the viewer's

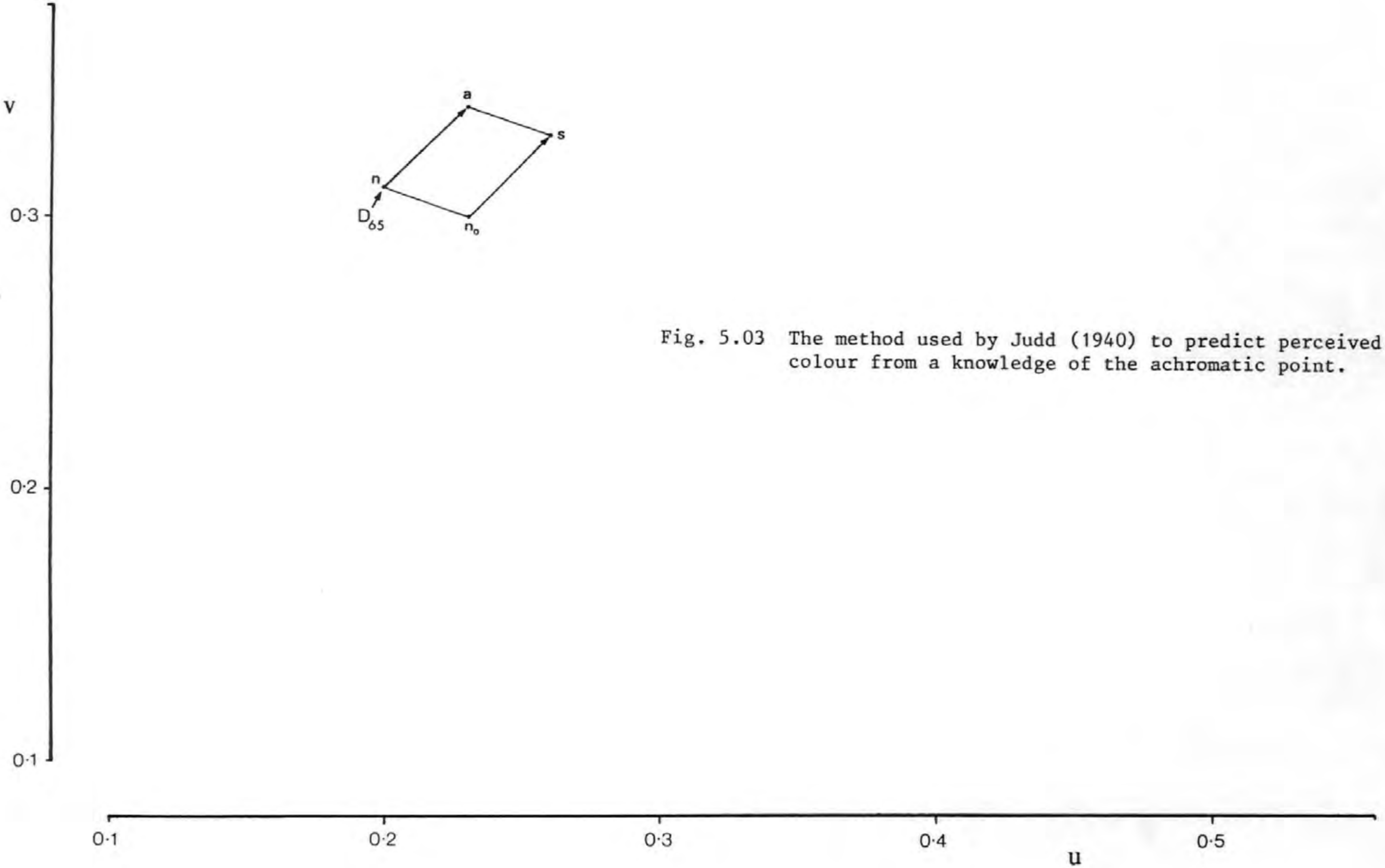


Fig. 5.03 The method used by Judd (1940) to predict perceived colour from a knowledge of the achromatic point.

state of chromatic adaptation is dominated by the display. If the perceived colour is to be predictable, therefore, the achromatic point must be known.

To measure the approximate position of the achromatic point for the above display, a foveal test area was generated, the chromaticity of which could be adjusted until the area appeared neutral. The system automatically maintained a constant ratio of test area luminance to surround luminance.

5.4 The system used for neutral point measurement.

The experimental test pattern is generated by the display system described in detail in chapters 3 and 4. A schematic of the experimental arrangement is given in Figure 5.04. A small square test area subtending approximately 1.5 degrees at the viewer's eye was positioned at the centre of the display. This area could be surrounded by various coloured stimuli as required. The luminance and chromaticity of the test area and the surround were set under computer control. The chromaticity of the test area could be changed quickly and easily by selected characters on the VDU keyboard, interpreted by the computer program as colour quantities. The program ensured that the luminance of the test area was kept constant and equal to that of the background as its chromaticity was changed.

To simplify test area colour adjustment, the VDU keyboard was masked so that only ten keys remained exposed as shown in Figure 5.05. The centre three keys were labelled red, green and blue whilst three keys to their left were -1, -5 and -10 units respectively. Three corresponding keys on the right were labelled +1, +5 and +10 units.

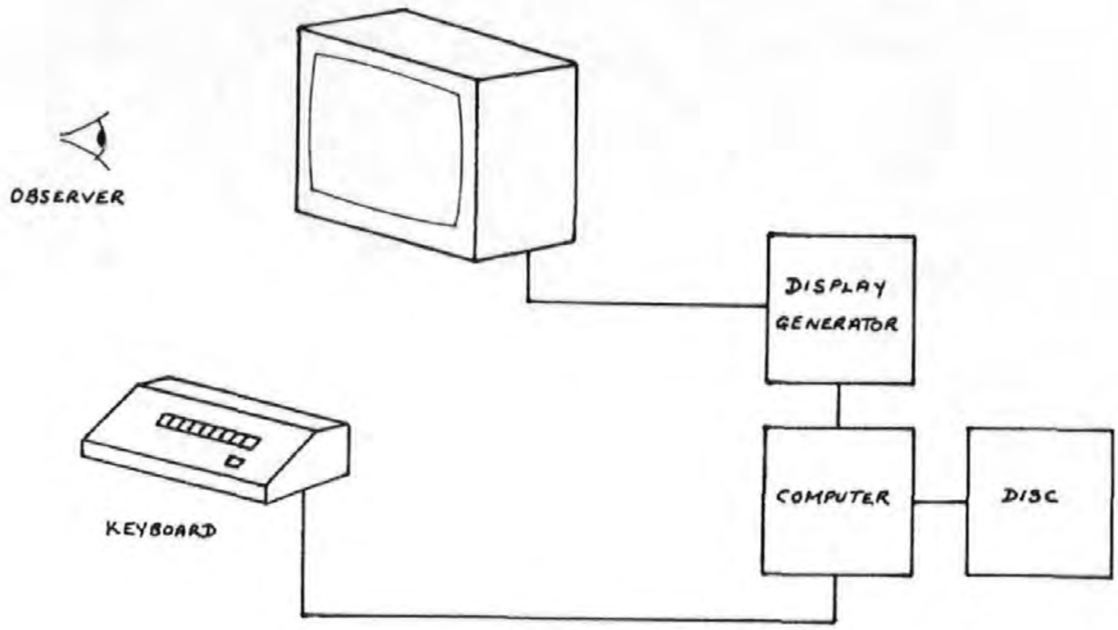


Fig. 5.04 Experimental arrangement for measurement of the achromatic point.

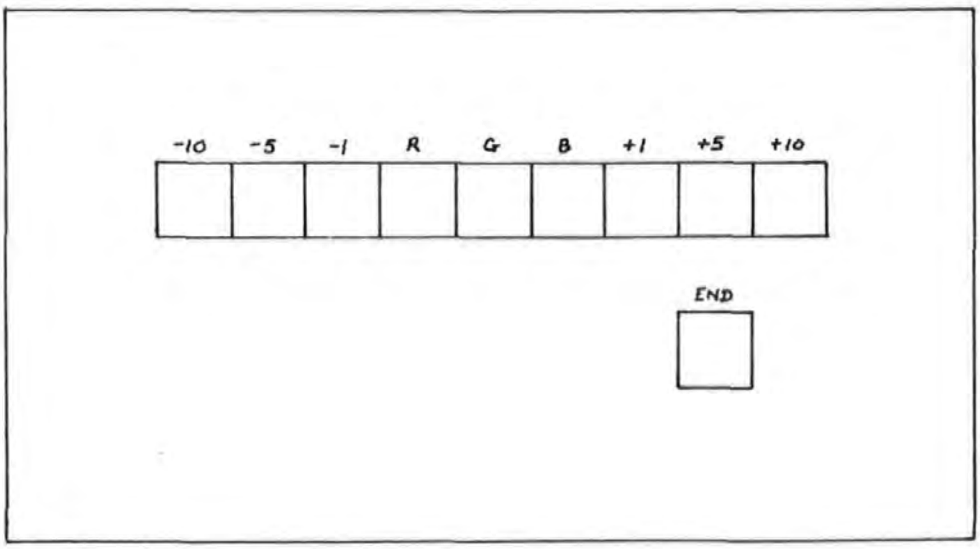


Fig. 5.05 Detail of the keyboard mask.

By selecting one of the three colour keys and then pressing one of the six quantity keys the relative amount of each 'primary' stimulus forming the test area could be varied, luminance remaining constant. Having selected the required test area colour, a tenth key was pressed to end the sequence and record the final chromaticity on magnetic disc.

As described in chapter 4, the display may be interfaced to the computer via an RS232 serial data link. This interface provided accurate data transmission over considerable distances, allowing the display to be used in a darkroom remotely from the PDP11-10 computer.

5.41 Experimental procedure.

All the measurements were carried out in a light-tight room with matte black surfaces. The display system was switched on approximately one hour before use to allow for thermal stabilisation. Using the computer, sample background colours were then set up on the display screen and the chromaticity and luminance measured using respectively a colorimeter and Hagner photometer. Any errors detected at this point would require the display monitor and display electronics to be checked for colour balance, convergence etc. as detailed in chapter 3. Given satisfactory chromaticity and luminance reproduction, the first background colour was displayed and the observer asked to view the display centrally for three minutes at a viewing distance of 45 cms. At the end of this period the test area was positioned at the screen centre, usually with a starting chromaticity matching that of illuminant D_{65}. The observer was asked to adjust the test area chromaticity until this appeared to be achromatic. Whilst adjusting the test colour, the observer was asked to fixate frequently outside the test area. When the test area had been adjusted to appear achromatic the observer pressed a key causing the corresponding chromaticity to be stored on magnetic disc and the next adapting stimulus to be displayed. The above procedure was then repeated.

Preliminary measurements were carried out using several colour-normal observers. A consistency of results was established amongst several subjects, thereafter the measurements use the author as the subject (sections 5.5 - 5.9).

5.42 The size of the test sample.

The area of the test sample must be small in order to minimise its direct adaptive effect. It is known that the size of a coloured stimulus also has a significant influence on colour discrimination (Brown, 1952[25]). To test the importance of test area size, the chromaticity of the test field perceived as neutral was measured for various test field areas with a fixed surround colour. Constant and equal luminances were maintained for test area and surround fields. The experiment showed, as might be expected, that the neutral shift increased as the area of the test patch decreased, see Figure 5.06. For a test area subtending 2 degrees or less, the shift in neutral point reached a limit, whilst colour discrimination was sufficient to produce consistent results for stimuli subtending 1 degree or more.

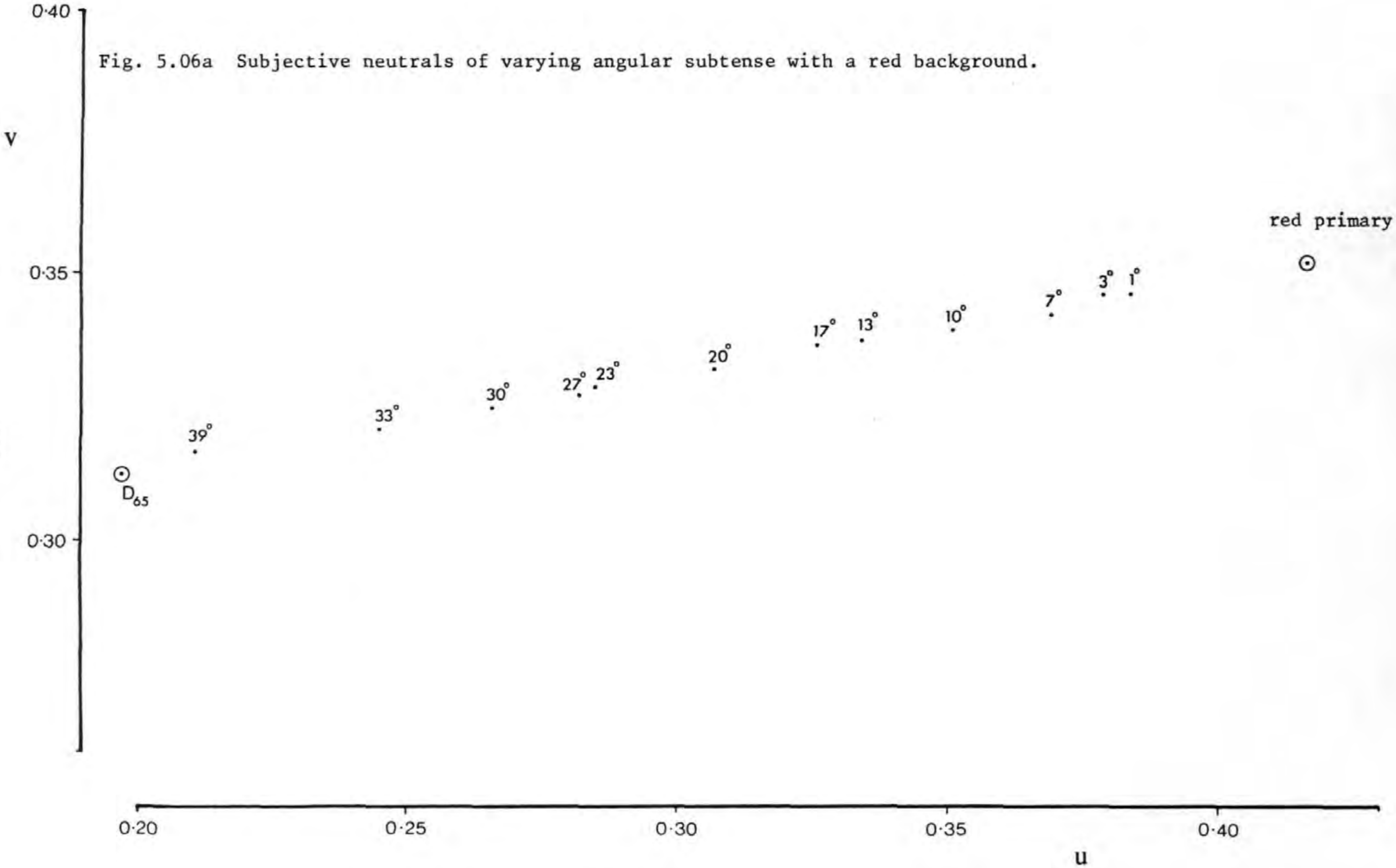


Fig. 5.06a Subjective neutrals of varying angular subtense with a red background.

Fig. 5.06b Subjective neutrals of varying angular subtense with a green background.

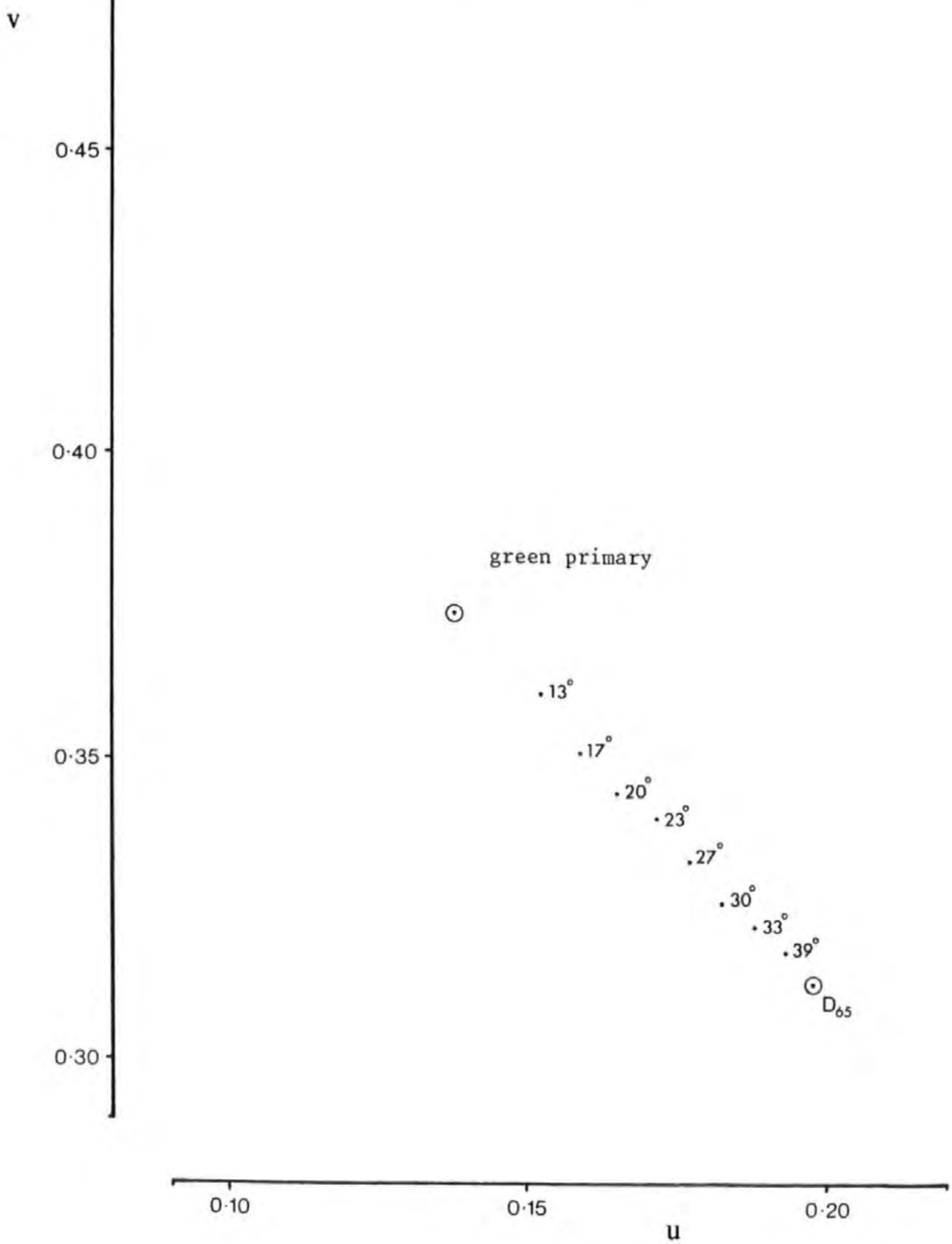
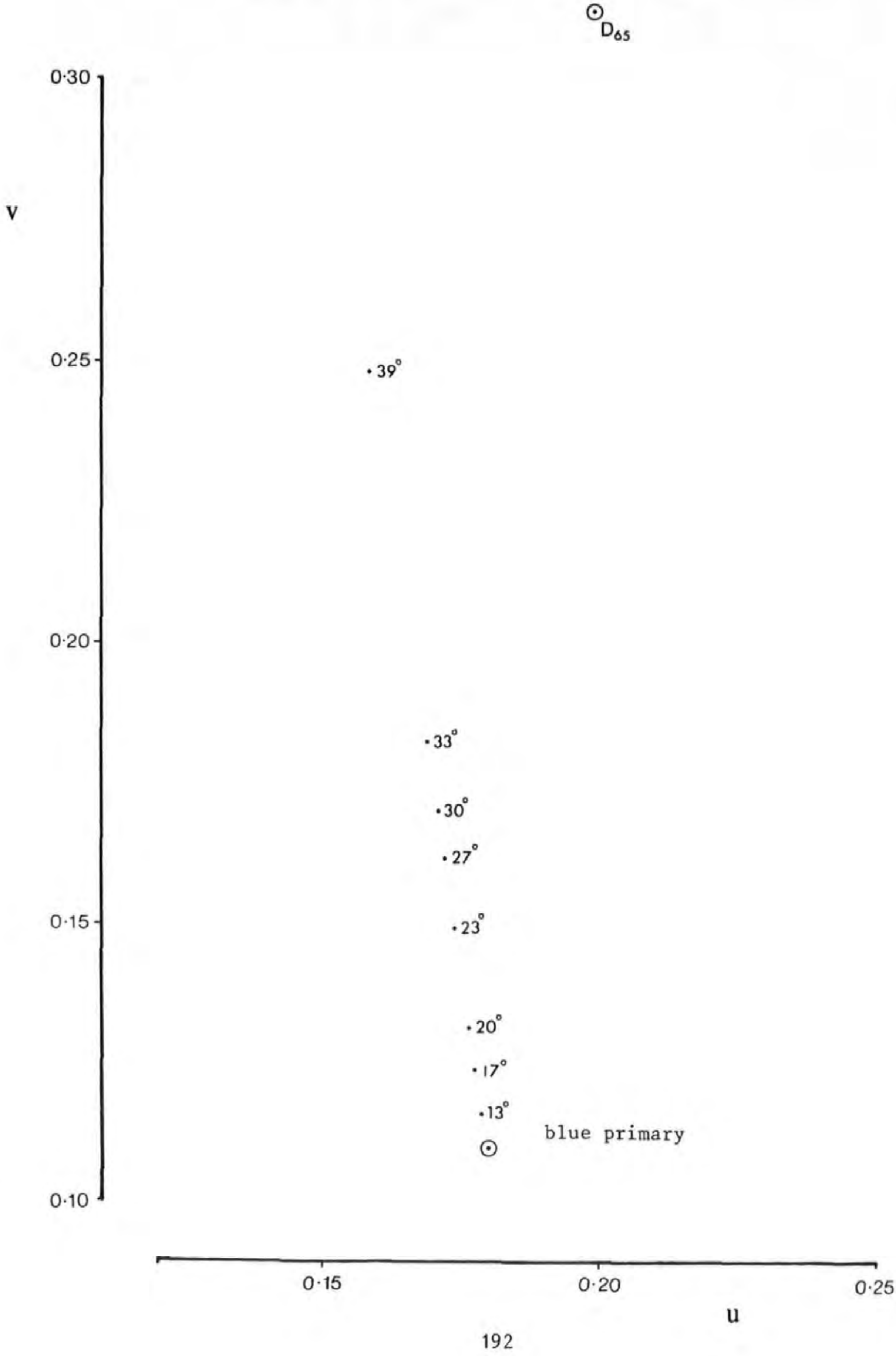


Fig. 5.06c Subjective neutrals of varying angular subtense with a blue background.



5.43 The role of local contrast in the test procedure.

The use of an adjustable neutral test area within a picture was described by Pearson (1975)[26] as an approximate measure of the achromatic point. Ideally, the test area should have no inductive or adaptive influence; in practice, its influence can only be minimised. MacAdam (1951)[27] showed that the colour shift produced by a large surround field or a small sample of the same luminance and chromaticity differed only in the magnitude of the colour shift when placed adjacent to a small test area; the perceived hue was the same in each case. The induction from the small chromatic field adjacent to and of the same size as the test area, was taken by MacAdam as a measure of the simultaneous contrast. The further colour shift, produced when a large area chromatic background was introduced, was used by MacAdam as a measure of chromatic adaptation.

To measure the effect of simultaneous contrast, the display system was arranged to display the adjustable test area adjacent to a fixed colour stimulus of the same area with a dark surround. Results of these measurements are shown in Figures 5.07 to 5.09, the chromatic stimuli being the chromaticities of the red, green and blue display phosphors in turn, with luminances respectively 75, 75 and 25 cd/m^2.

These achromatic points were qualitatively consistent with the results of MacAdam, showing that a shift of the stimulus perceived as neutral does occur in this case. This shift was greatest for the blue stimulus, the neutral lying close to the Munsell hue locus intercepting the chromaticity of the blue phosphor. The green stimulus produced much less colour shift, whilst that produced by the red phosphor was a distance of only 0.014 in the CIE UCS diagram. The important influence of blue

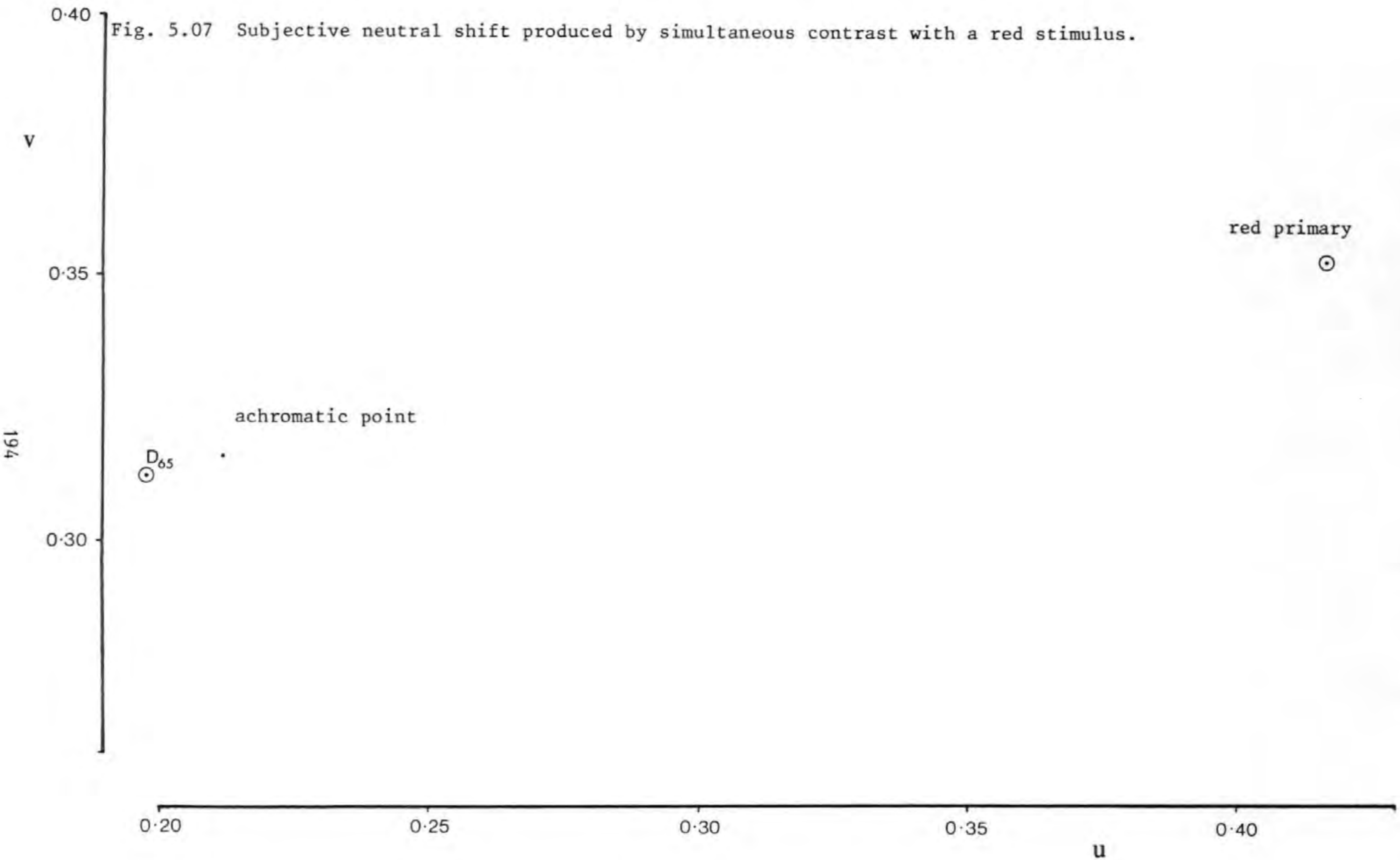


Fig. 5.07 Subjective neutral shift produced by simultaneous contrast with a red stimulus.

Fig. 5.08 Subjective neutral shift produced by simultaneous contrast with a green stimulus.

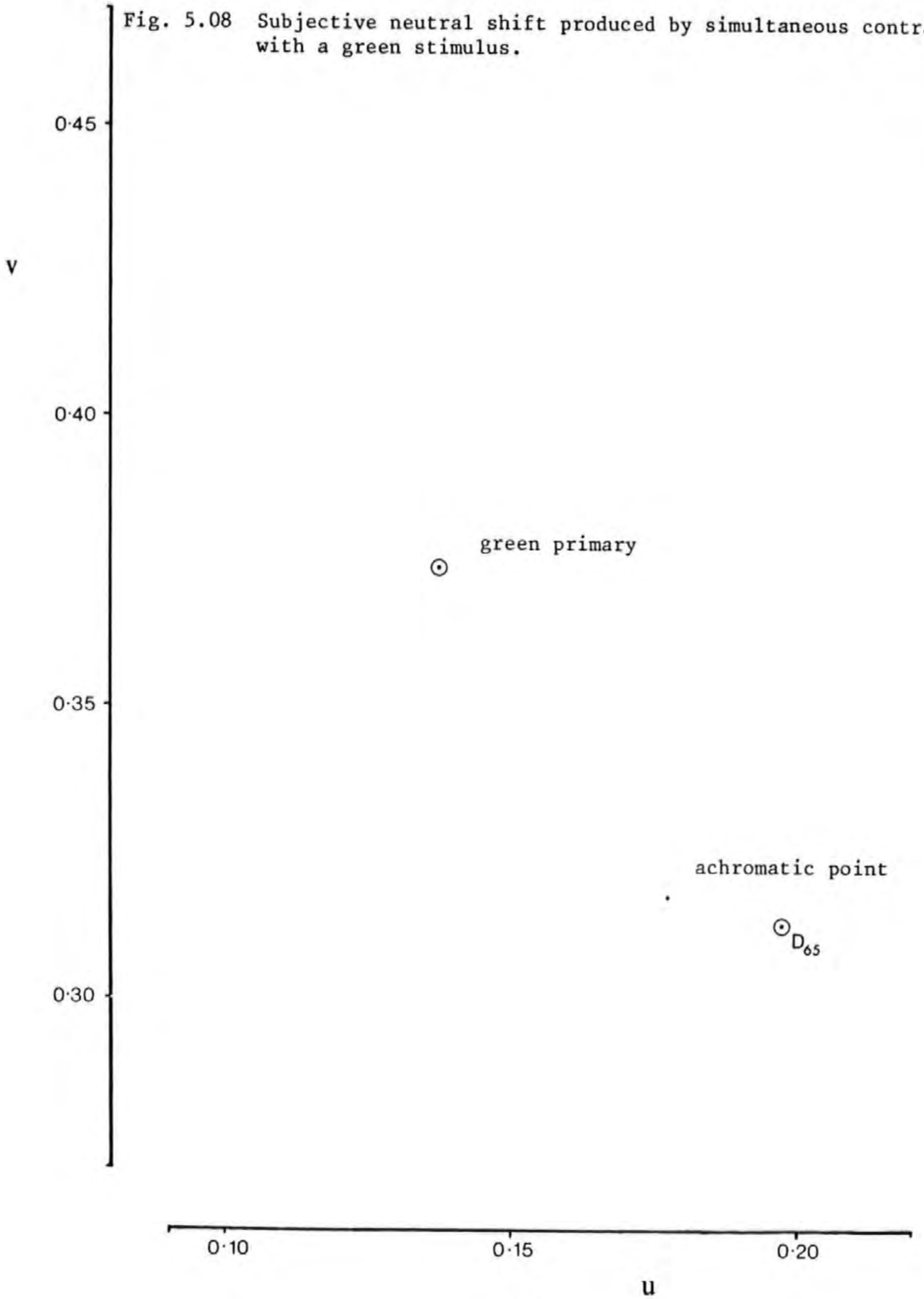
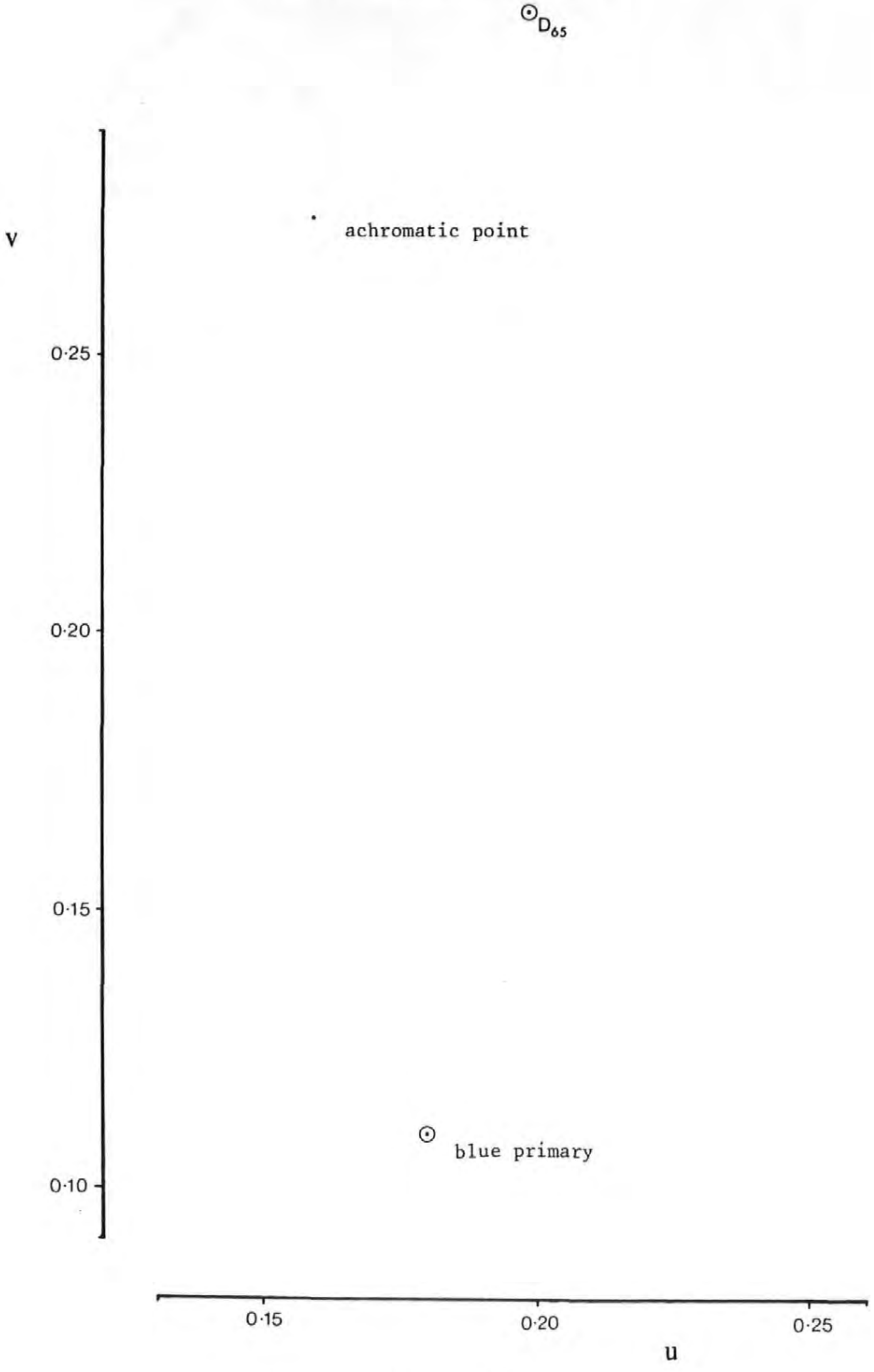


Fig. 5.09 Subjective neutral shift produced by simultaneous contrast with a blue stimulus.



stimuli has been observed elsewhere (Helson and Michels, 1948[28]; MacAdam, 1950[29]; Sugimoto et al, 1973[8]; Kinney, 1967[30]) and is further discussed at the end of this chapter. The purpose of this experiment was to show the quantitative effect of simultaneous contrast as defined by MacAdam so that its effect could be compared with that of a large surround area as measured in later experiments.

In their study of adaptive effects when viewing pictures, both from projected transparencies and on television screens, Hunt and Winter (1975)[21] avoided simultaneous contrast or induction by using a test area outside the main picture area, using a colour naming approach to find the achromatic point. This method assumes that, during the time required for the fixation point to switch to the test area, the level of adaptation remains unchanged. As a comparison of the experimental methods, a sequence of background colours was set up on the display screen, matching in chromaticity and luminance some of the colours used by Hunt and Winter. The results of this experiment, given in Figure 5.10, show several features.

1. Large shifts in the chromaticity of the adapting stimulus produce correspondingly smaller but significant changes in the chromaticity perceived as neutral, always in the direction of the adapting stimulus.

2. The achromatic points follow approximately the constant hue loci.

3. For background chromaticities with colour temperatures close to 6500°K, adaptation is almost complete.

These results are in qualitative agreement with those of Hunt and Winter shown in Figure 5.11.

Fig. 5.10 Subjective neutrals measured from the display screen for background chromaticities matching those of Hunt and Winter.

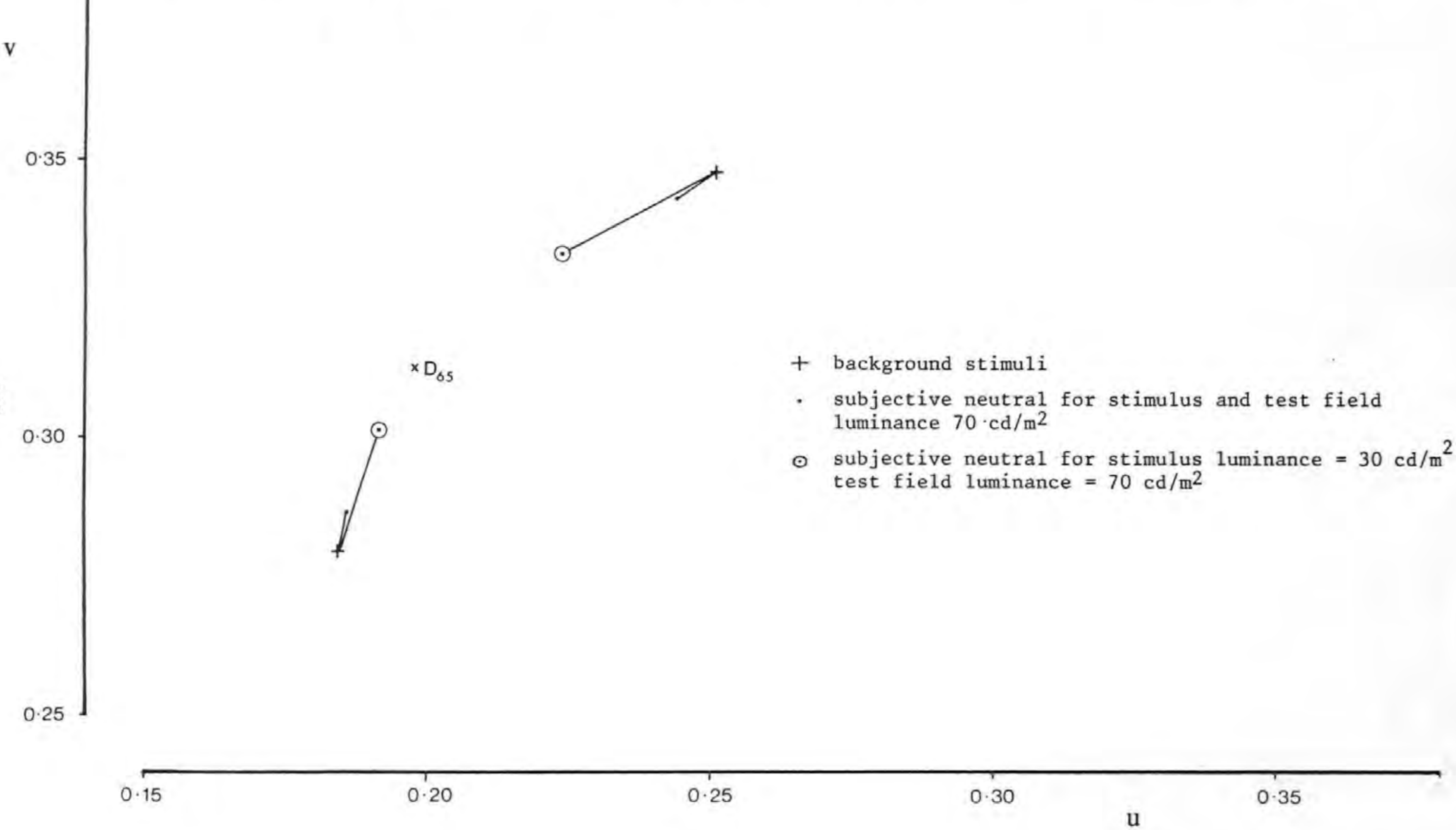
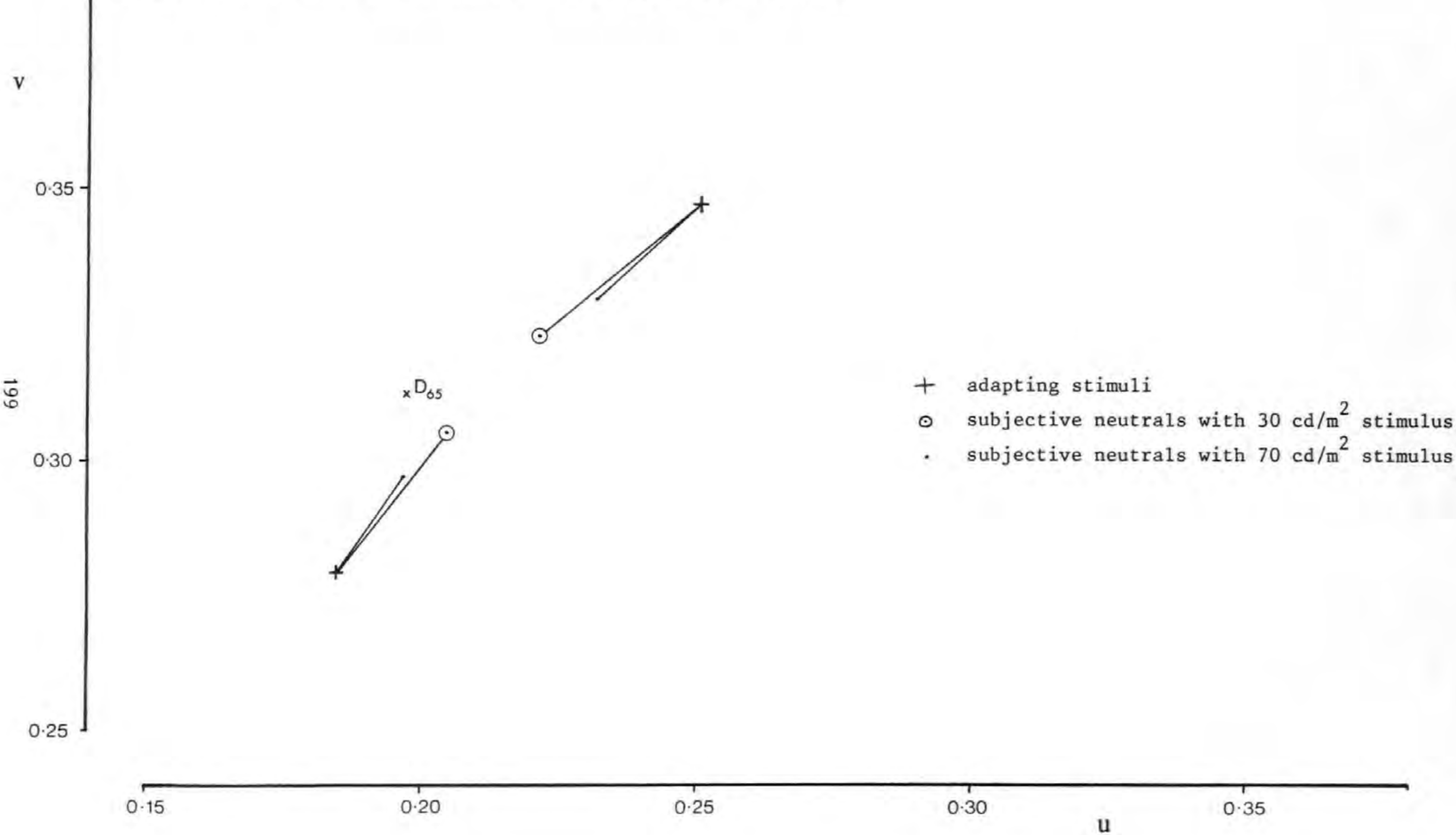


Fig 5.11 Subjective neutral points measured by Hunt and Winter.



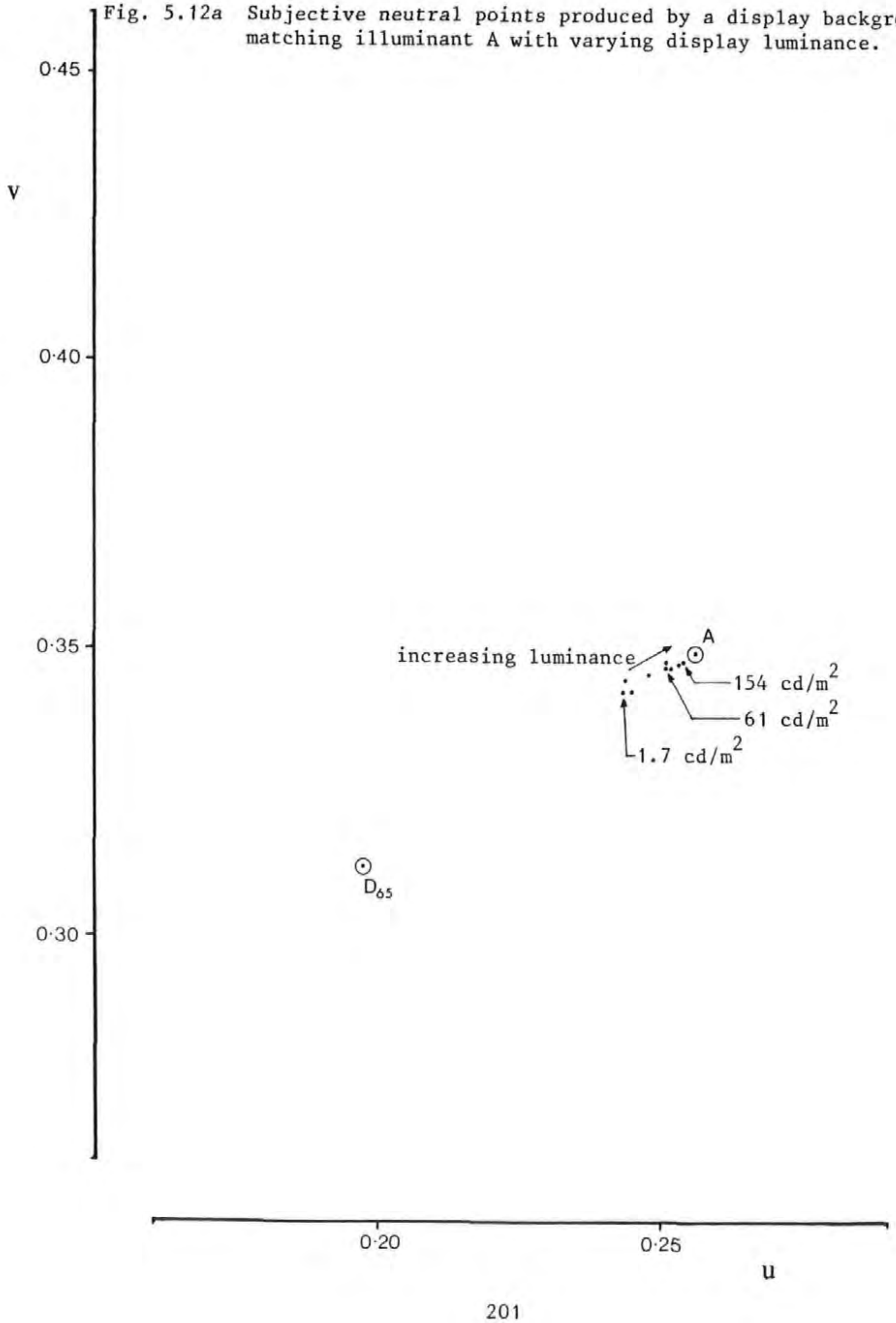
The achromatic points produced by the display system show a greater displacement from the daylight point than those of Hunt and Winter. This increased chromaticity shift would be anticipated since the measurements from the display system included induction and involved a background stimulus subtending 45° compared to the 30° stimuli used by Hunt and Winter. It was felt that the achromatic point should be representative of colour appearance within the display, and therefore appropriate that the effects of induction should be included in its measurement.

5.5 The achromatic point produced by a display chromaticity matching illuminant A. Some observations on the effect of luminance.

During the work described in chapter 3, it was noted that, for the special case of a wholely 'white' environment, in which all the room surfaces match the chromaticity of D_{65}, changing the simulated illuminant produced large changes of the perceived surface colour. In all cases, with an illuminant chromaticity matching D_{65}, the surfaces were perceived as neutral. If, however, an illuminant matching the chromaticity of illuminant A was selected, the screen was always perceived to have a 'yellow' tint.

Figure 5.12 shows the chromaticity of the test area perceived as neutral for a display screen uniformly lit with the chromaticity of illuminant A (u = 0.256, v = 0.349). Several neutral points are shown, indicating the observed neutral shift as the display luminance was increased. The background and test area luminances were equal. In all cases the neutral point shows a significant shift towards the chromaticity of the stimulus. This shift is seen to increase as the screen luminance

Fig. 5.12a Subjective neutral points produced by a display background matching illuminant A with varying display luminance.



is increased. The neutral shift, even at maximum luminance (150 cd/m^2), is always incomplete. Over a typical display luminance range (10 - 70 cd/m^2) the magnitude of the neutral point chromaticity shift is relatively small.

The shift in the neutral point chromaticity towards that of the background, as the whole screen luminance is increased, is consistent with the observation by Hurvich and Jameson (1957)[31] that a wide range of colour temperatures is capable of producing a 'white' sensation, given a high enough luminance level. The results given in Figure 5.12 show that as luminance increases, the perceived background colour becomes less saturated implying increased chromatic adaptation.

5.6 The effect on the achromatic point produced by varying saturation.

To study the effect of varying the saturation of the colour stimulus, three sets of chromaticities were produced along straight lines in the CIE UCS diagram with equal chromaticity steps between the reference neutral, D_{65}, and each of the display primary chromaticities. In each case the luminance was set close to the maximum for each primary, 75 cd/m^2 for red and green, 25 cd/m^2 for blue. The results of section 5.5 indicate that lower luminances will produce a small reduction in adaptive shift. The display screen was uniformly lit with each of the colour sequences in turn. The results are shown in Figures 5.13a, b and c.

As found in section 5.5, shifts in the neutral point chromaticity were produced by all the colour stimuli, these being in the direction of the chromaticity of each stimulus. As the saturation of the stimulus

Fig. 5.13a Subjective neutrals produced by red backgrounds of varying saturation.

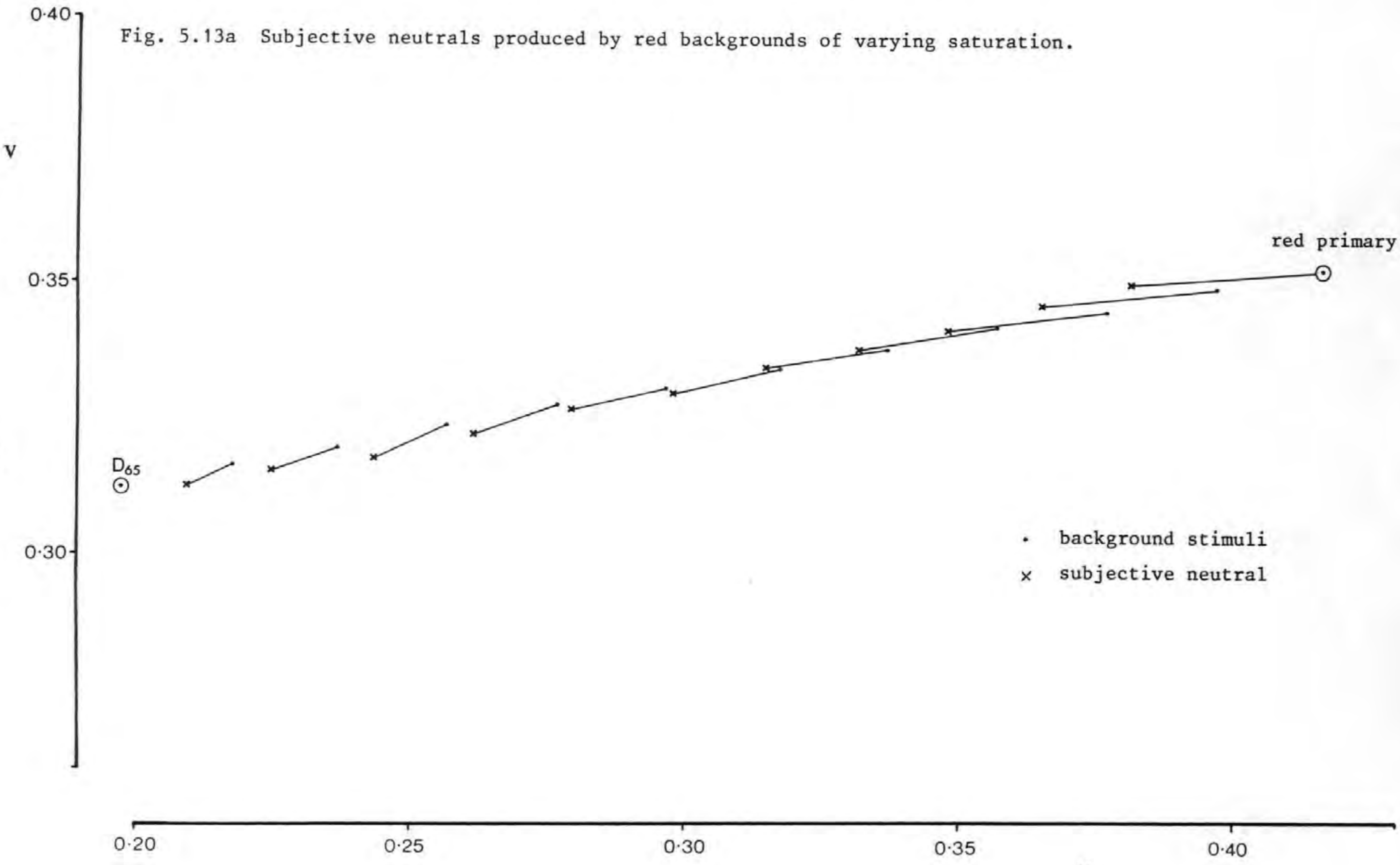


Fig. 5.13b Subjective neutrals produced by green backgrounds of varying saturation.

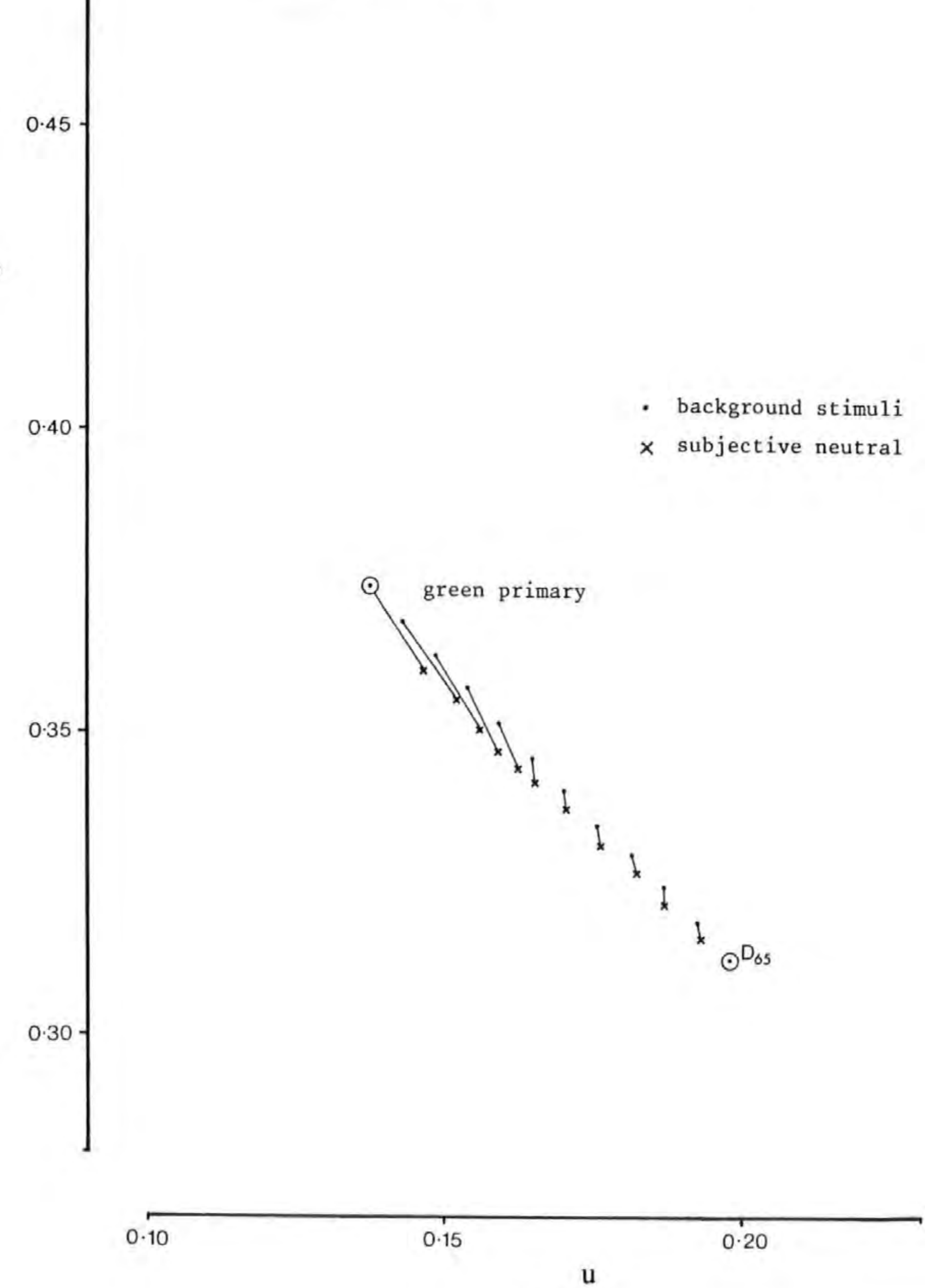
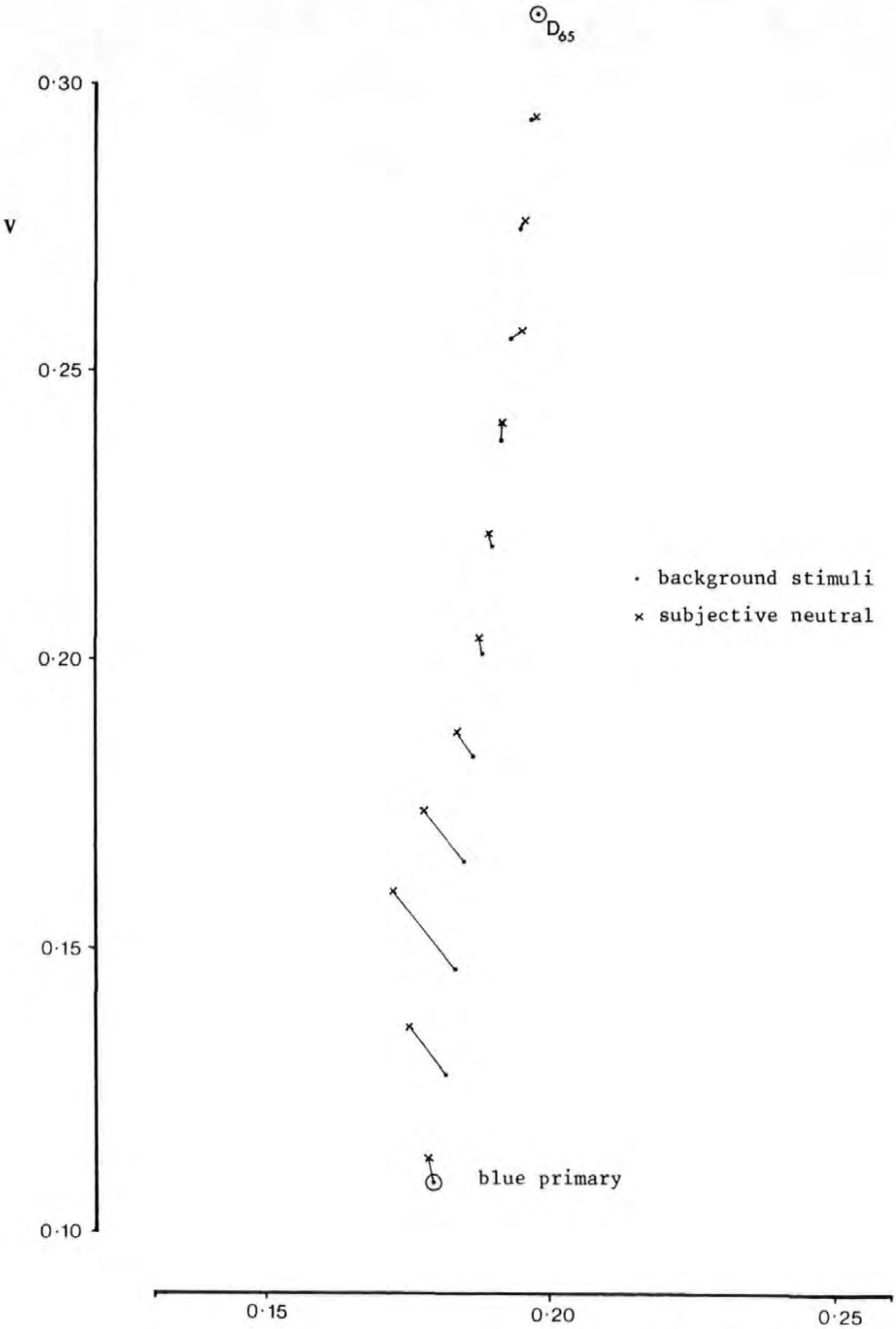


Fig.5.13c Subjective neutrals produced by blue backgrounds of varying saturation.



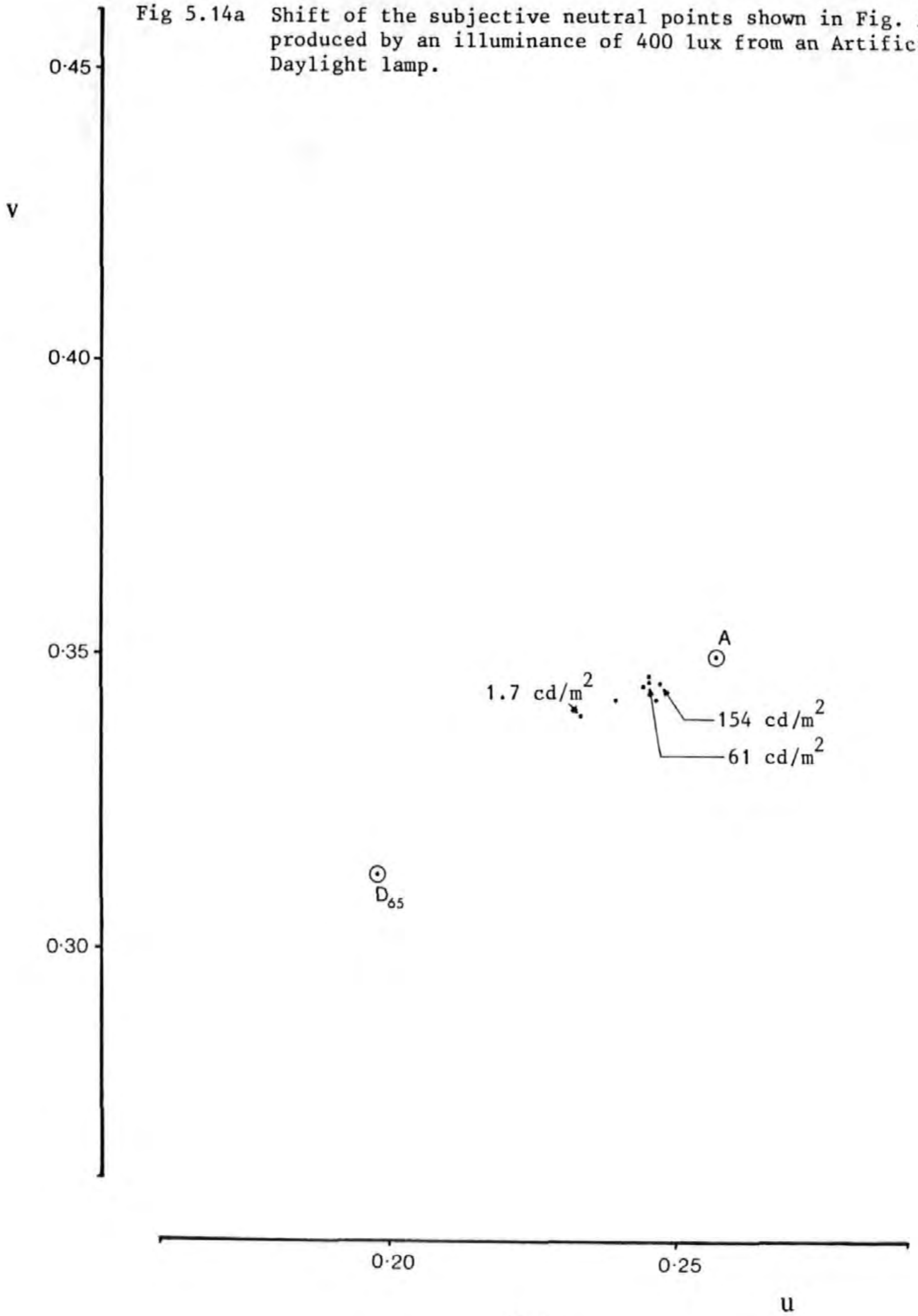
increases, the chromaticity of the neutral point also moves away from the central area of the display colour gamut. The level of adaptation to the blue display phosphor is, in general, higher than that produced by the green phosphor, which itself produces more shift in the neutral point than the red phosphor.

5.7 The influence of surround illumination on the perceived neutral.

In chapter 3, the effect of ambient light scattered from the display screen was shown by its objective effect on the measured display chromaticity. To obtain some measure of the effect on the perceived white point, the experiment described in section 5.5, measuring the adaptation to the chromaticity of illuminant A at various display luminance levels, was repeated with ambient lighting from two Artificial Daylight fluorescent lamps, producing an illumination of 400 lux in the vertical plane of the display at the screen viewer's eye level. The results of this experiment are shown in Figure 5.14.

The neutral points show a shift towards the stimulus similar to that produced without surround lighting, but in this case all the neutral points are moved slightly towards the chromaticity of the ambient light source. This shift indicates that even at a display luminance of 150 cd/m^2, a surround illuminance of 400 lux produces a measurable change in the achromatic point. It should be emphasized that the achromatic points were measured for an image subtending at 45° with a dark surround. The perceived colour shift is consistent with the scattering of light from the display screen rather than peripheral adaptation.

Fig 5.14a Shift of the subjective neutral points shown in Fig. 5.12 produced by an illuminance of 400 lux from an Artificial Daylight lamp.



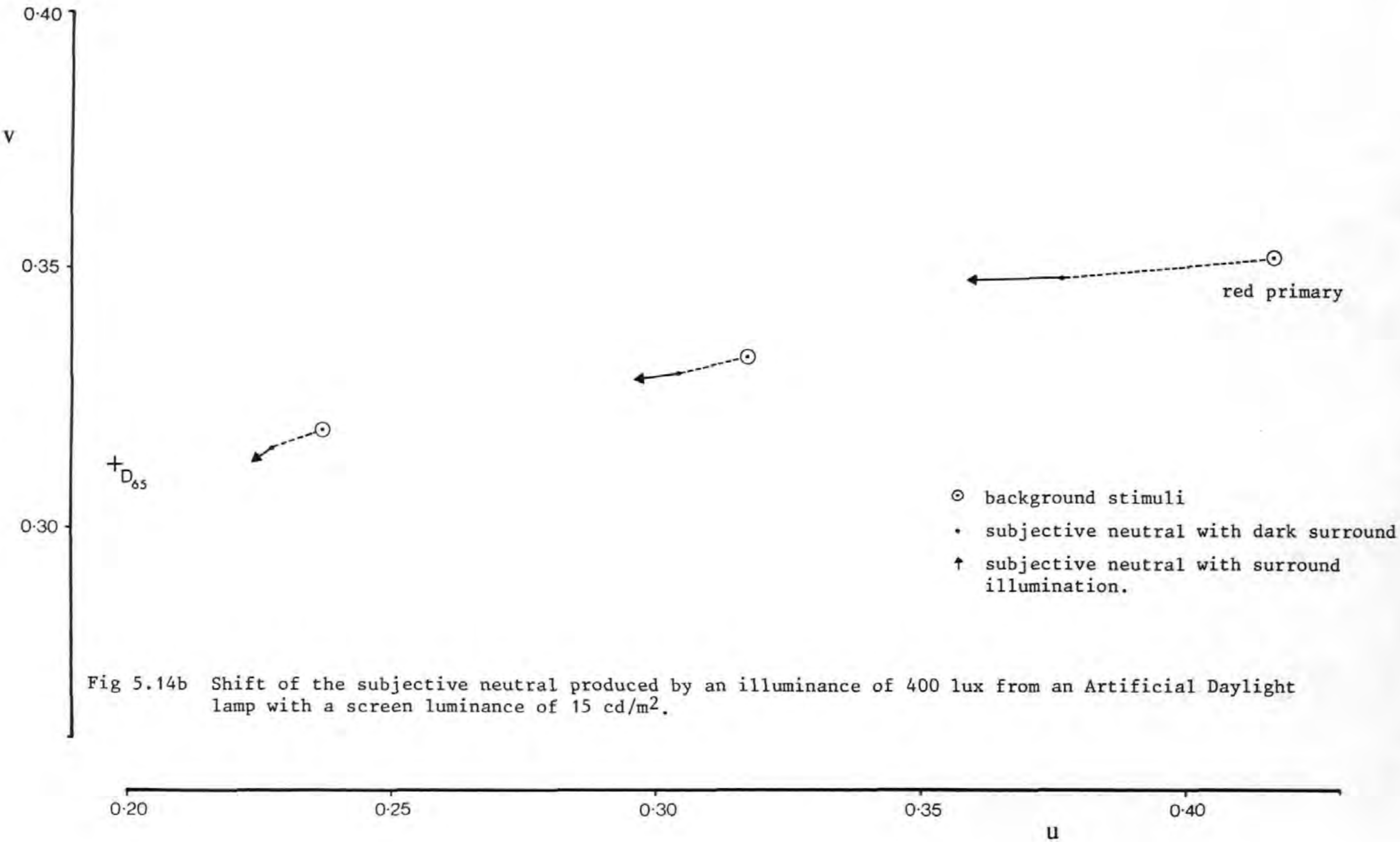


Fig 5.14b Shift of the subjective neutral produced by an illuminance of 400 lux from an Artificial Daylight lamp with a screen luminance of 15 cd/m^2.

Fig. 5.14c Shift of the subjective neutral produced by an illuminance of 400 lux from an Artificial Daylight lamp with a screen luminance of 70 cd/m^2.

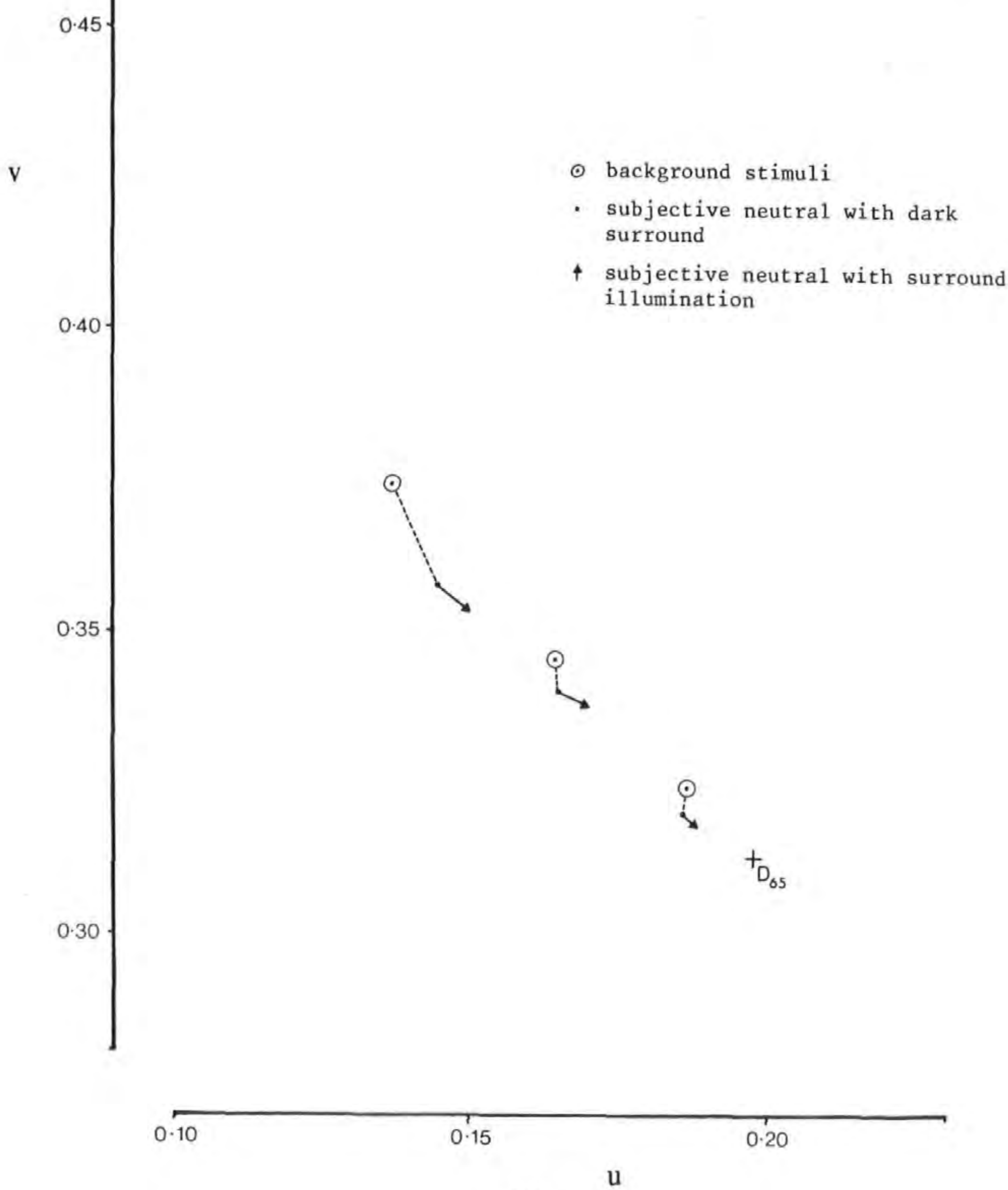
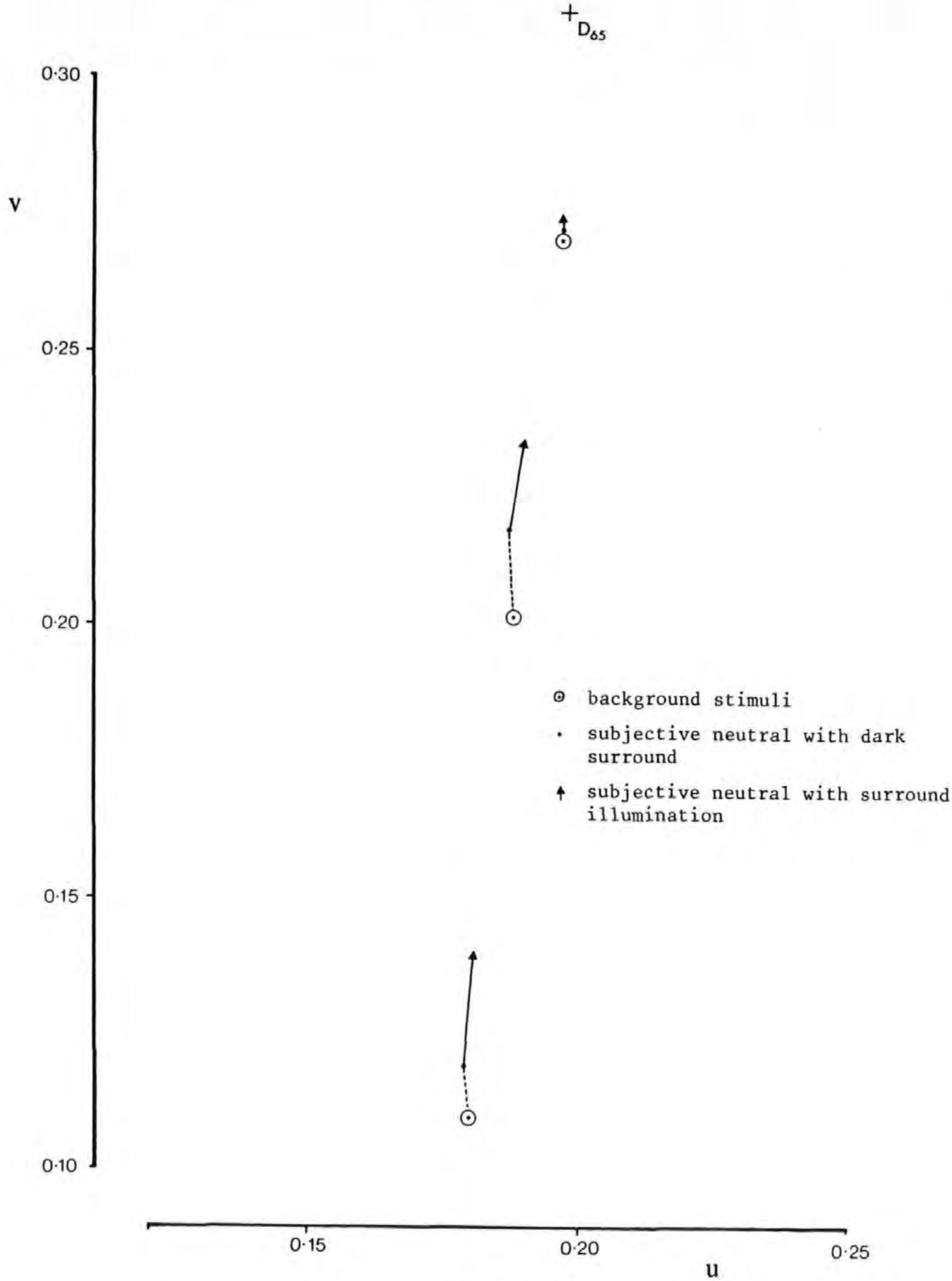


Fig.5.14d Shift of the subjective neutral produced by an illuminance of 400 lux from an Artificial Daylight lamp with a screen luminance of 25 cd/m^2.



5.8 The effect of luminance contrast on the perceived neutral.

The work by Land (1959)[32] and Pearson et al (1969)[33] and Richards and Parks (1971)[22] on the measurement and prediction of perceived colour suggested that the luminance of a colour stimulus relative to its surroundings could have a strong influence on the resulting colour perception. Hunt and Winter (1975)[21] measured the neutral point for a number of stimuli with chromaticities along the black-body locus and showed that if the adapting luminance is doubled relative to that of the test area, a considerable shift in the chromaticity perceived as neutral should be expected.

To examine the effect of relative luminance, a display program was written which allowed for any background chromaticity within the display gamut to be selected, following which a series of test areas was generated, each with a different luminance relative to the fixed background. The observer was asked to find the achromatic point for each luminance ratio using the procedure described in section 5.41.

The results of this experiment are shown in Figure 5.15 for several background chromaticities. It is clear that the luminance of the test area relative to the adapting stimulus has a significant effect on the chromaticity perceived as neutral. Increase of the test area luminance relative to that of the background produces a shift of the neutral point away from that of the background. Conversely, reduction of the test area luminance below that of the background moves the neutral point closer to the chromaticity of the background; in this case the visibility of the test area is impaired and colour discrimination becomes increasingly difficult.

Fig. 5.15a Subjective neutral points for varying luminance ratio with a red background of luminance 75 cd/m^2.

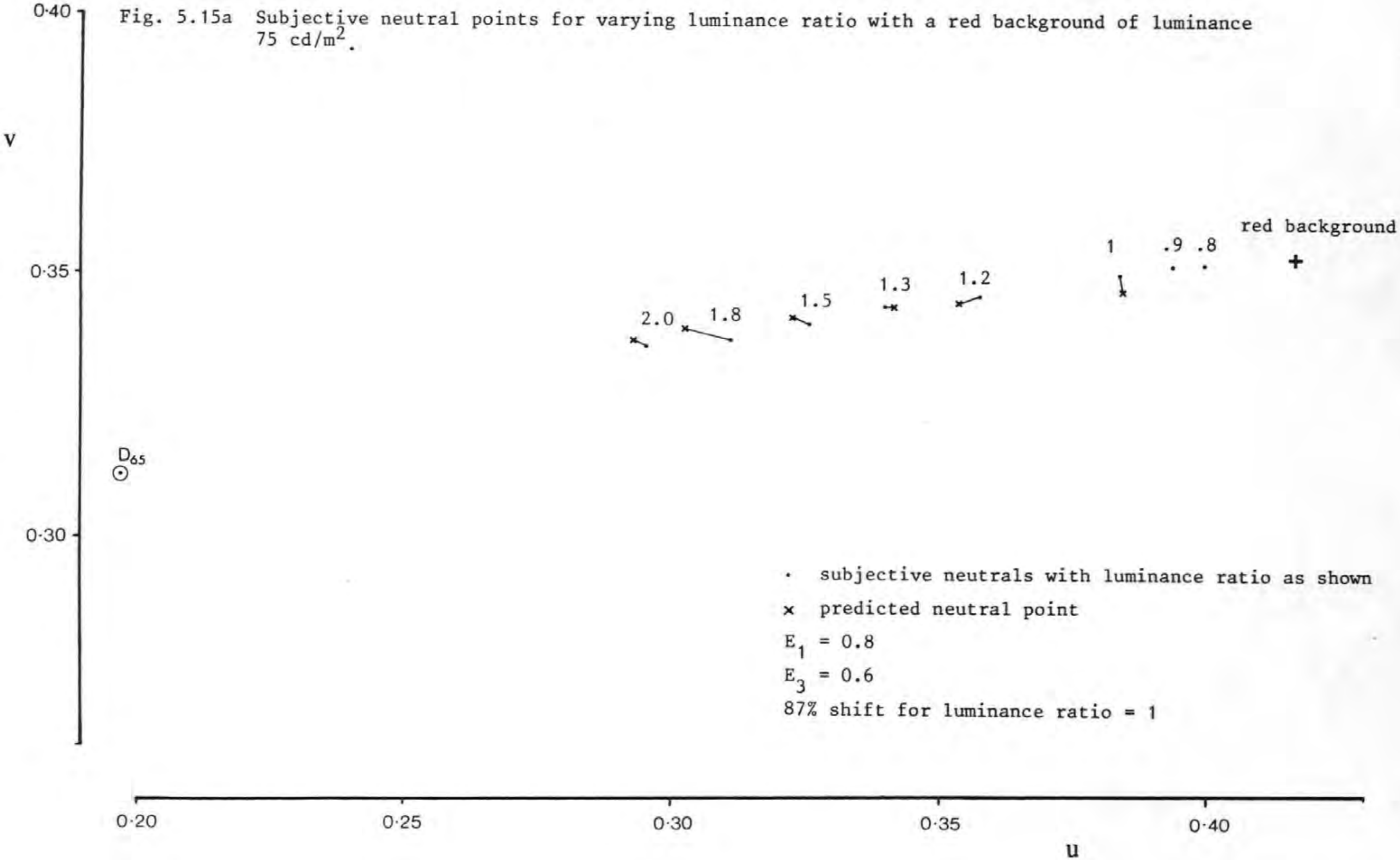


Fig. 5.15b Subjective neutral points for varying luminance ratio with a green background of luminance 75 cd/m^2.

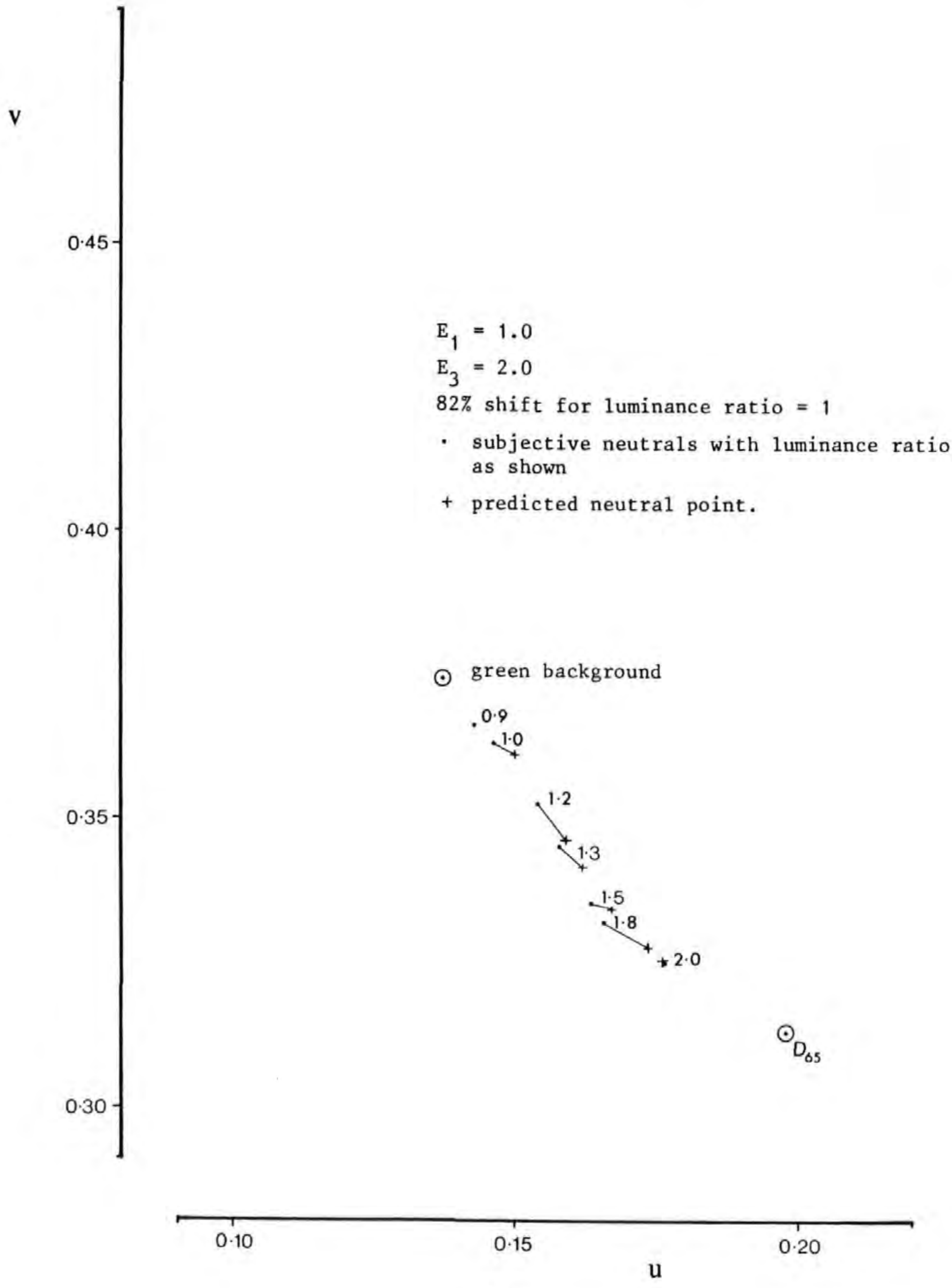


Fig. 5.15c Subjective neutral points for varying luminance ratio ratio with a blue background of luminance 25 cd/m^2.

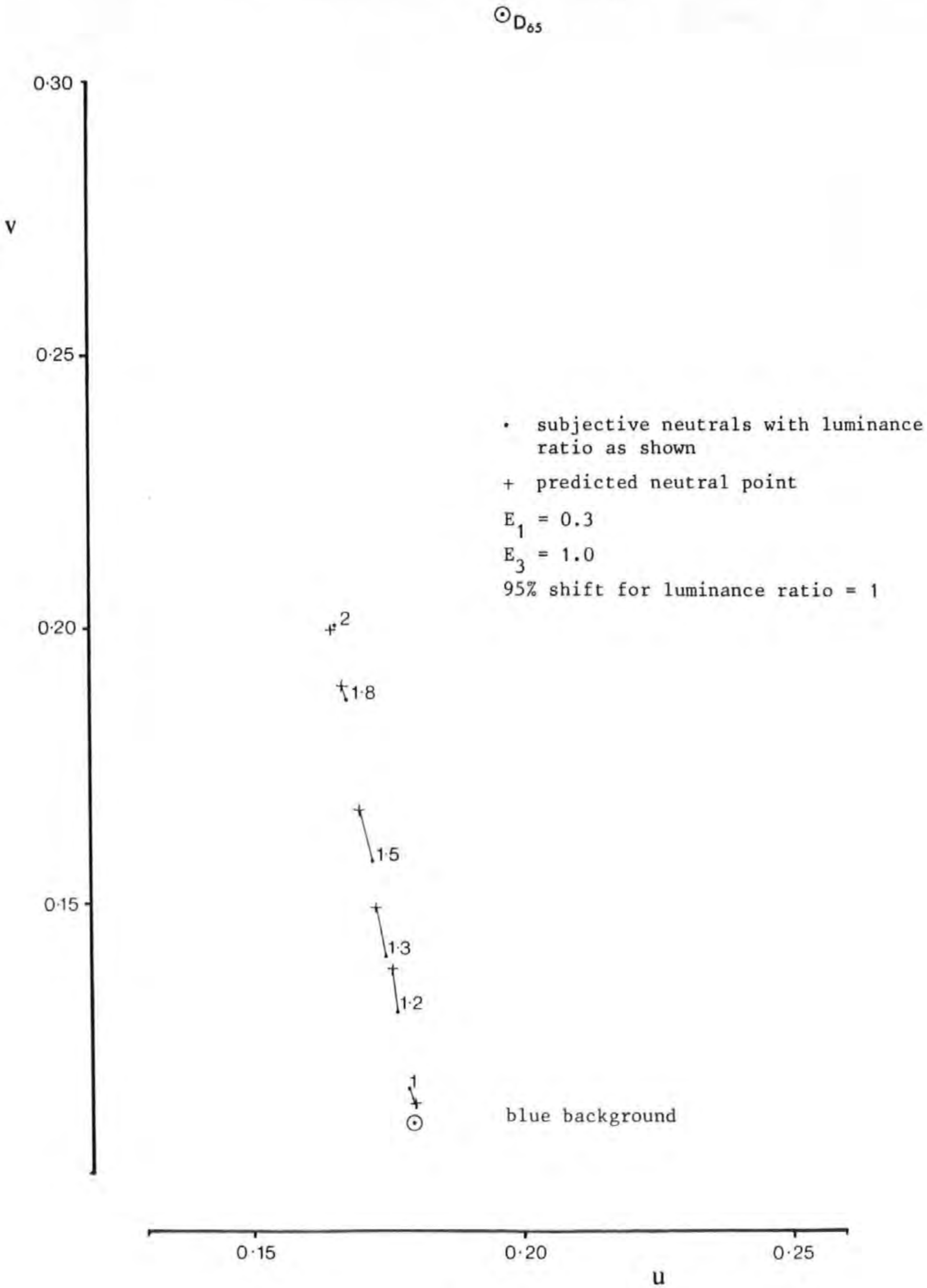


Fig. 5.15d Subjective neutral points for varying luminance ratio for a background colour matching that of a 'Grolux' fluorescent tube of luminance 40 cd/m^2.

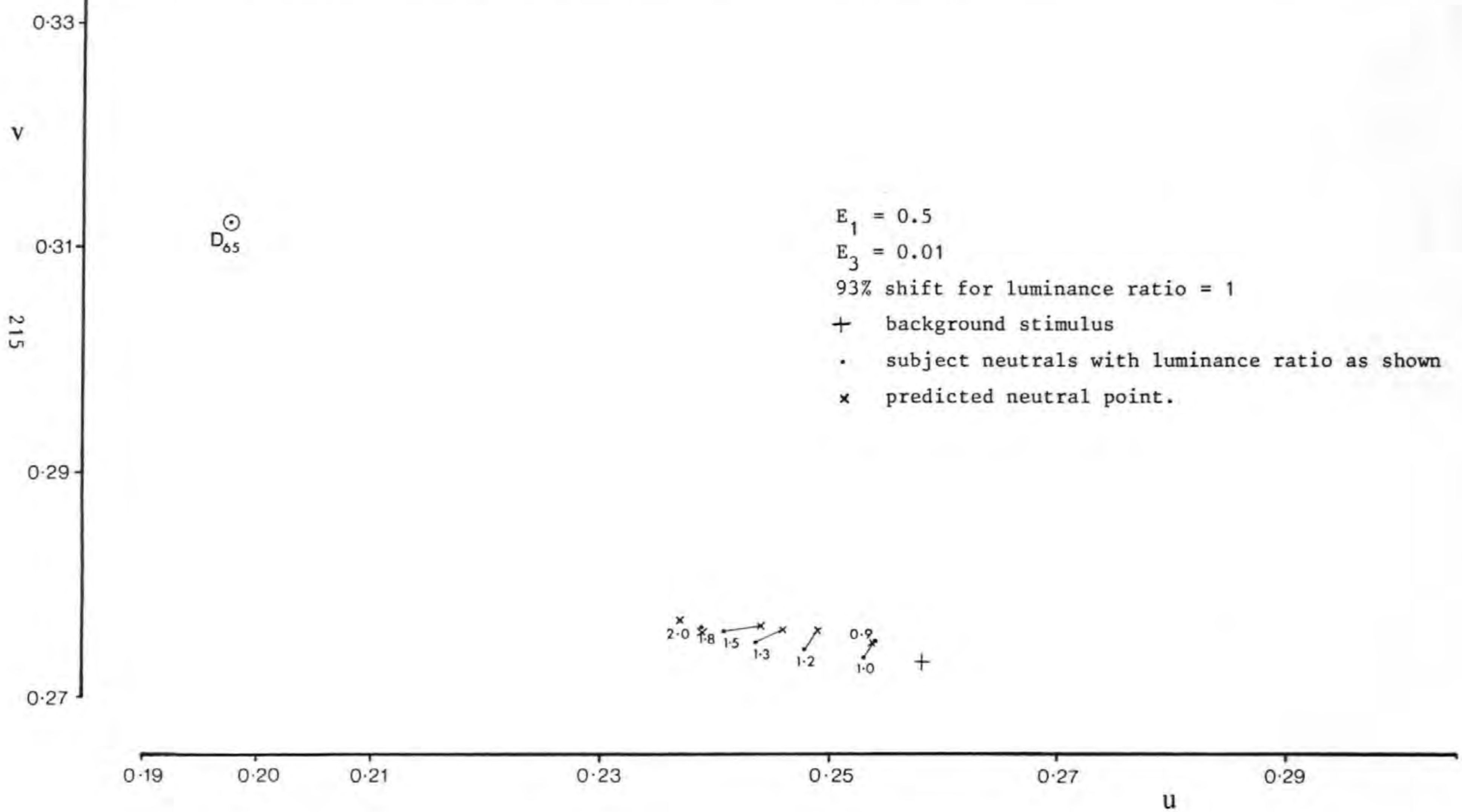
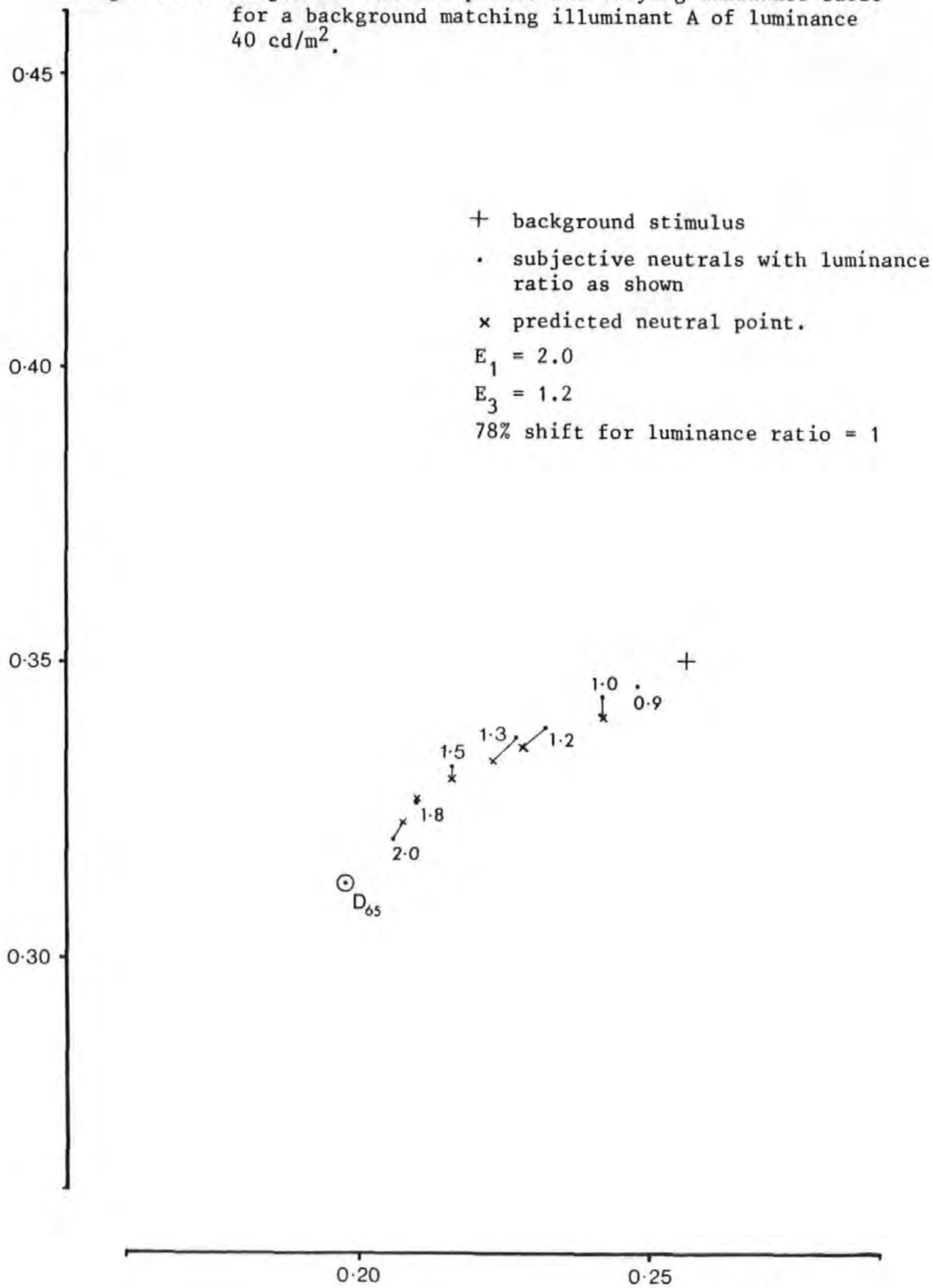


Fig. 5.15e Subjective neutral points for varying luminance ratio for a background matching illuminant A of luminance 40 cd/m^2.



The measured achromatic points form curves in the u,v chromaticity diagram as the luminance ratio is changed, suggesting that responses may be non-linear. The results add further support to the suggestion by Richards and Parks (1971)[22] that a different achromatic point is appropriate for each level of luminance.

5.9 The effect of multiple stimuli on the achromatic point.

The achromatic point measurements described above are confined to the case of a single colour background. In order to measure the achromatic point for a multi-coloured image, the simple arrangement shown in Figure 5.16 was used. Several pairs of equal-luminance colours were displayed, the luminance being fixed at 25 cd/m^2 to allow the use of the full chromaticity gamut of the display. Approximate achromatic points were measured using the procedure described in section 5.4. The results of these measurements are shown in Figure 5.17a-c.

These measurements indicate that where the colour balance point is close to the reference neutral chromaticity, the achromatic point also lies close to this chromaticity. When an off-balance colour combination is used, producing a colour balance point some distance from the reference neutral, the perceived neutral is also displaced. In this case, if the two colour stimuli have a large hue difference and high saturation, the displacement of the achromatic point does not relate in a straightforward manner to the position of the colour balance point, this being particularly noticeable for red-green colour pairs. There is a tendency for the achromatic points to be biassed towards the blue-green boundary of the chromaticity gamut.

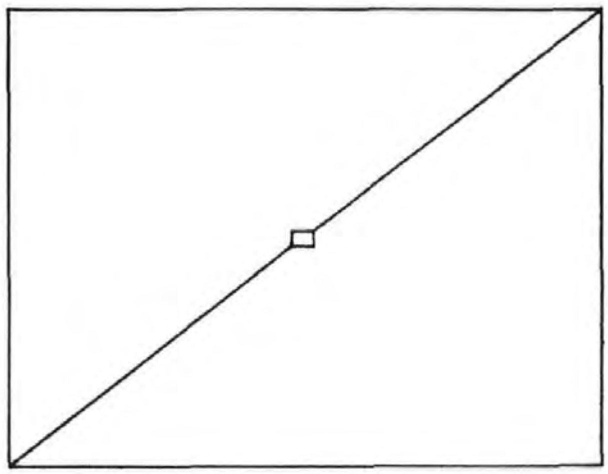


Fig. 5.16 The form of the image for measuring the subjective neutral produced by two colours.

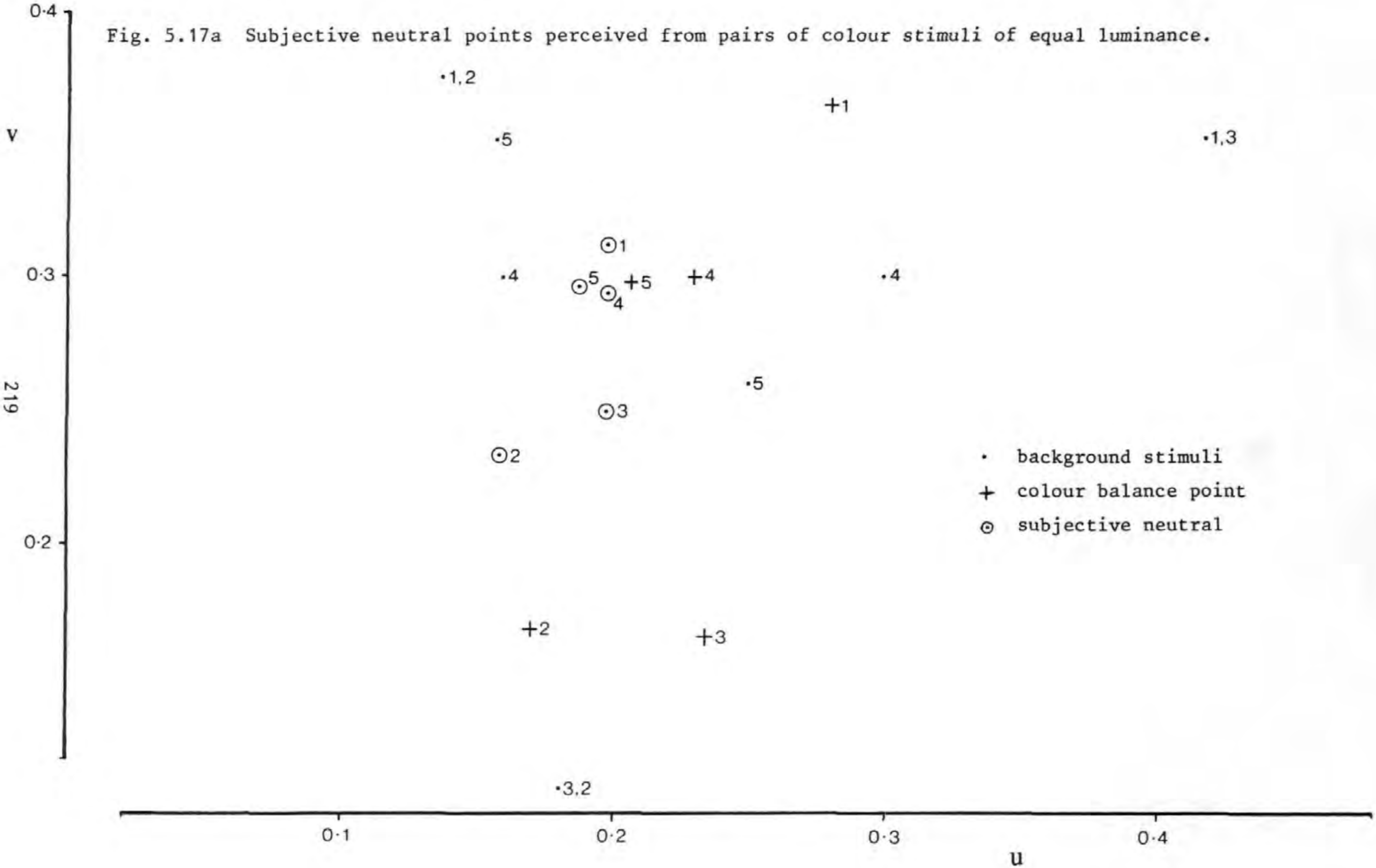


Fig. 5.17a Subjective neutral points perceived from pairs of colour stimuli of equal luminance.

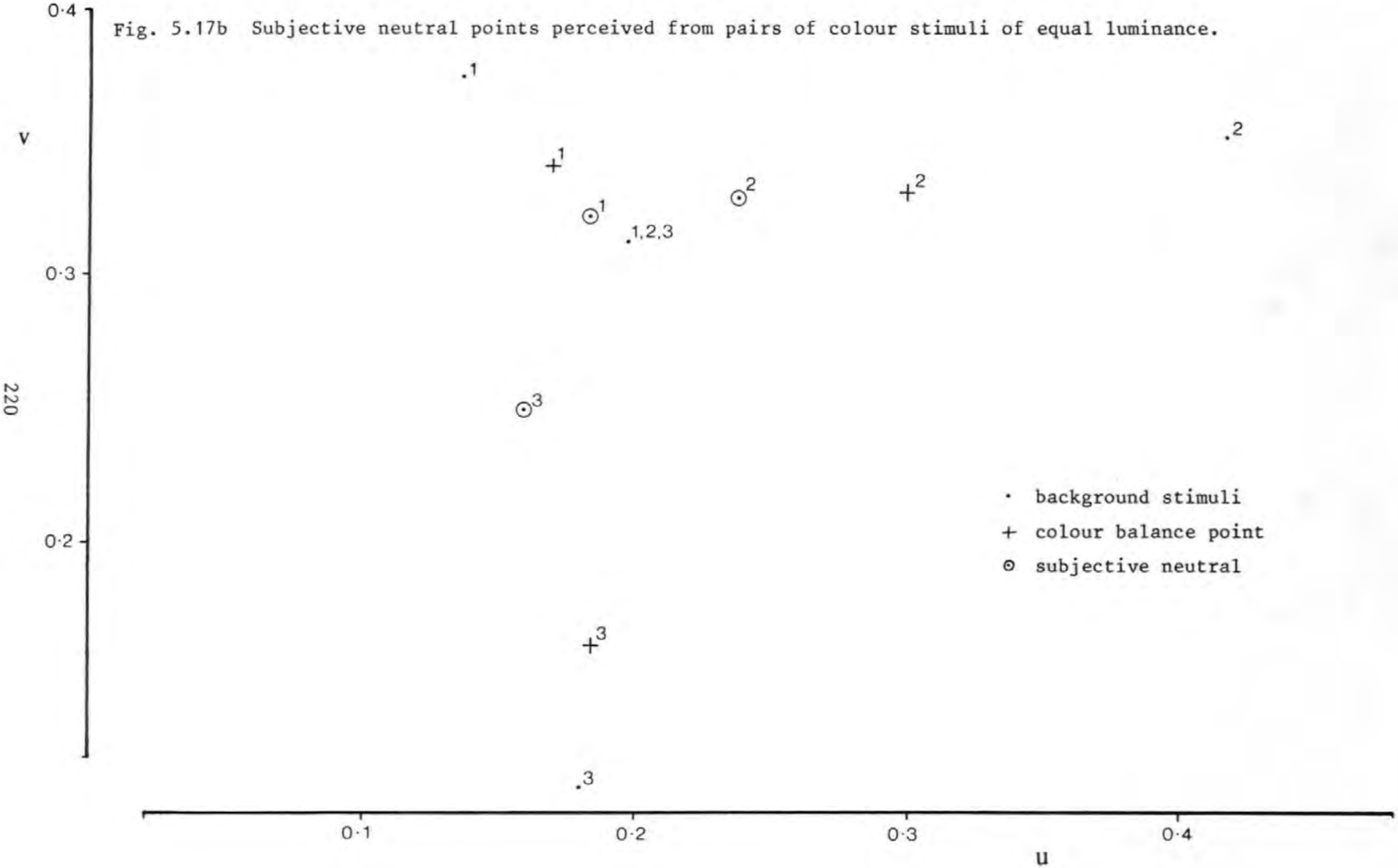
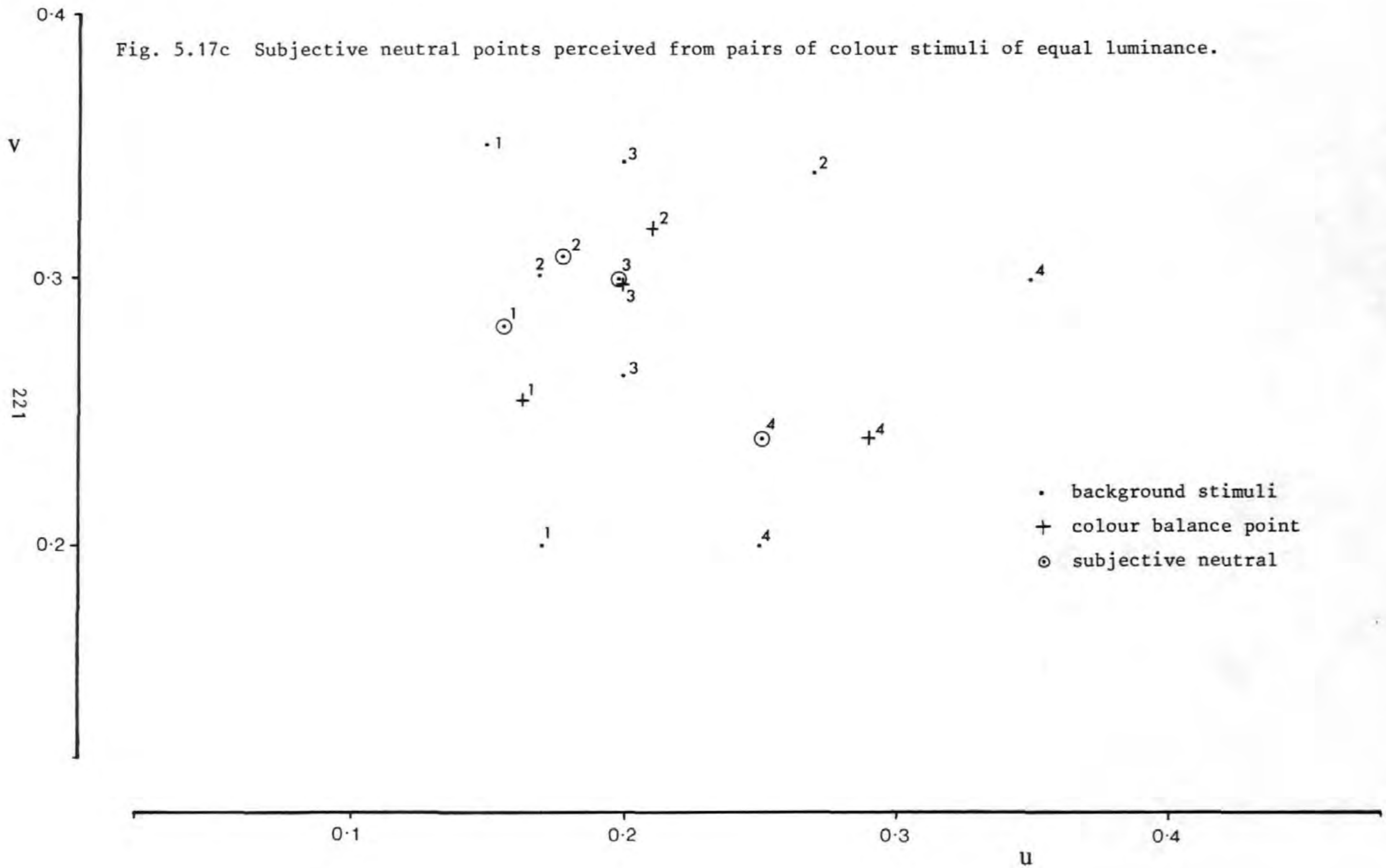


Fig. 5.17b Subjective neutral points perceived from pairs of colour stimuli of equal luminance.

Fig. 5.17c Subjective neutral points perceived from pairs of colour stimuli of equal luminance.



When half the display area matches the reference neutral, the displacement of the colour balance point is restricted, ensuring that the achromatic point, as shown in Figure 5.17b, remains relatively close to the reference neutral.

5.10 Prediction of the neutral point.

The results described in sections 5.5 to 5.9 serve to emphasize that a colour stimulus of a given chromaticity can take on a variety of colour appearances, depending on its environment.

Judd (1940)[20] published a model based on empirical equations which purported to predict colours perceived in complex images. Interest in this model was renewed in 1959 when Land demonstrated his two-colour separation overlay images. Judd (1960)[34] suggested that the Land phenomena were consistent with established theory and could be predicted from his 1940 model. In 1969, Pearson et al[33] confirmed that the Judd equations were able to predict the Land perceptions with a high level of accuracy. A simpler and mathematically more elegant model was developed by Richards and Parks (1971)[22]. Like the Judd model, this too was shown to predict the Land perceptions. Both these models are based on the prediction of the achromatic point. Having established the chromaticity giving neutral perception, the colours perceived from other stimuli within the image may be predicted. These models are discussed more fully in Appendix 5.

Richards and Parks also suggested that the model might be capable of predicting other chromatic induction effects such as those demonstrated by Kinney (1967)[30]. For this reason it was decided to investigate the possibility of using the Richards-Parks model to predict achromatic perception for the colour display.

The prediction was based on manipulation of the three fundamental responses used by Richards and Parks:-

$$R(\lambda) = 0.6\bar{x}(\lambda) + 0.5\bar{y}(\lambda) - 0.1\bar{z}(\lambda)$$

$$Y(\lambda) = \bar{y}(\lambda)$$

$$B(\lambda) = 0.67\bar{y}(\lambda) + 0.33\bar{z}(\lambda)$$

These functions had been found to give a best fit to MacAdam's chromatic adaptation results.

In terms of the C.I.E. uniform colour space

$$\therefore \quad R = 0.75U + 0.8V - 0.2W$$

$$Y = V$$

$$B = 0.495U - 0.32V + 0.66W$$

The first step in the prediction procedure represents 'discounting the illuminant'. The reference white point is taken to be the chromaticity u = 0.1978, v = 0.3122 matching illuminant D_{65} with the same luminance as the background stimulus. This stimulus chromaticity and the reference white point are expressed in terms of the fundamental responses giving $R_1Y_1B_1$ and $R_2Y_2B_2$ respectively. The 'discounting level' is expressed as the percentage shift P of the achromatic point between the reference white and the chromaticity of the illuminant. This may be expressed in terms of von Kries coefficients K_i (i = r,y,b) as described in Appendix 5 , where:

$$K_r = 1/((R_1P/100R_2) + (1 - P/100))$$

$$K_y = 1$$

$$K_b = 1/((B_1P/100B_2) + (1 - P100))$$

where the luminance of the achromatic point is assumed equal to the average for the visual field = L_{av}.

Taking account of luminance contrast, when L_s is the luminance of the achromatic area and L_{av} is the average background luminance, the von Kries coefficients are modified so that

$$K_i' = K_i(L_{av}/L_s)^{E_i} \qquad i = r,y,b$$

where E_i is a constant for each response, dependent on the viewing conditions.

Using the modified von Kries coefficients, the fundamental response levels corresponding to the achromatic point were given by

$$R_3 = R_2 \times \frac{L_s}{L_{av}} \times K_r'$$

$$Y_3 = L_s$$

$$B_3 = B_2 \times \frac{L_s}{L_{av}} \times K_B'$$

Finally, the responses $R_3Y_3B_3$ were transformed back to the CIE UCS tristimulus values using the matrix:

$$U = 1.11R - 0.78Y + 0.34B$$

$$V = Y$$

$$W = -0.83R + 1.07Y + 1.26B$$

The values U,V,W define the chromaticity of the stimulus which would appear achromatic for a luminance level L_s. Having derived the values K_i' for a particular level of luminance, the perceived chromaticity, i.e. the chromaticity of a matching aperture stimulus, may be predicted for any other stimulus chromaticity.

The outstanding problem concerns the values to be assigned to the constants E_i. These were derived empirically to give a best fit to the measured achromatic points. Appropriate values are shown in Figure 5.15.

The results shown in Figure 5.15 indicate that the two-stage non-linear von Kries model can predict the chromaticity of the achromatic stimulus for a simple background-target image, as the target luminance is varied. Support is also shown for the suggestion by Richards and Parks[22] that the model would give successful predictions even when highly chromatic stimuli are used.

In addition to predicting the achromatic stimulus for a given background chromaticity, the model can predict the perceived colour for other stimulus chromaticities, assuming their luminance is known, and that the achromatic point remains stable. Conversely, if the achromatic point is known for the average luminance level, it is possible, in principle, to determine the chromaticity required in the display to produce a particular colour sensation.

The Richards-Parks model proved successful for predicting the perceived colours in a two-dimensional display of coloured patches[22]. It therefore seems reasonable that the model could be applied to the television display described in chapter 3. Suppose that the colours perceived in a real room are known in terms of the chromaticities of matching aperture stimuli. The colour balance point $\bar{u},\bar{v}$ is first calculated for these chromaticities, weighted according to luminance and area. This chromaticity is defined by the mixture of the three average luminances L_R, L_G and L_B for each display primary using the colour mixture formula described in chapter 3

$$\text{where} \quad L_R = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} L_{R_{ij}}$$

$$L_G = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} L_{G_{ij}}$$

$$L_B = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} L_{B_{ij}}$$

where $L_{R_{ij}}$, $L_{G_{ij}}$ and $L_{B_{ij}}$ are the respective luminances of the red, green and blue display primaries for picture element ij, part of an MxN array.

If the chromaticity $\bar{u},\bar{v}$ is close to the display reference neutral chromaticity, the achromatic point is assumed to coincide with this chromaticity and the colours are displayed as colorimetric matches for the colours perceived in the real room. If, however, the colour balance point differs from the reference neutral point, the average-luminance achromatic point is calculated, based on the appropriate percentage shift from the reference neutral to the colour balance point. The chromaticities of the surface colours are next modified, to give the correct hue and saturation relative to the achromatic point appropriate to each level of luminance, by using the vector procedure adopted by Judd[20]. A new colour balance point is now calculated, based on these modified chromaticities, from which a new achromatic point is predicted, the process being repeated until the required level of convergence is achieved. The final set of surface chromaticities are transferred to the display screen.

This procedure is proposed as a basis for further study based, possibly, on a colour naming approach to perceived colour measurement.

5.11 Discussion.

When presented with an unorthodox image, such as the room interior simulation described in chapter 3, the viewer makes a choice between three possibilities in order to arrive at a final perception. The image consists of either: (1) a two-dimensional pattern of coloured patches; (2) a scene under daylight-white illumination with surface colours essentially those perceived for each viewed in isolation; or (3) a scene under tinted illumination with perceived surface colours modified in some way so as to discount the illuminant colour. A detailed study of the mechanisms which govern the perception of colour is outside the scope of this thesis; a review of some relevant aspects of the subject is given in Appendix 5 .

The series of experiments described in the earlier sections of this chapter represents a preliminary investigation of colour perception from a display screen. These experiments involved very simple viewing conditions and the results are regarded as a basis for further study rather than a final answer to the problem of controlling colour perception. The results are summarised:-

Measurements described in section 5.5 demonstrated that a colour stimulus with chromaticity close to that of illuminant D_{65}, viewed in isolation, produces an immediate and stable neutral sensation for the dark-adapted observer. This observation is consistent with the results of Hurvich and Jameson (1951)[12] and Hunt and Winter (1975)[21]. As the chromaticity of the dominant stimulus is changed, the chromaticity producing a neutral sensation also changes. This achromatic shift is always in the direction of, but is smaller than, that of the stimulus chromaticity. This finding is in qualitative agreement with many studies of chromatic adaptation, notably those of MacAdam[27,29] and Hunt and Winter[21].

If the luminance of the display is increased, for both the background and test stimuli equally, a small shift in the achromatic point occurs, again in the direction of the background chromaticity. The increased luminance appears to produce an increase in the level of chromatic adaptation and hence a desaturation of the perceived colour. This result is perhaps not surprising, since Hurvich and Jameson (1951)[12] have shown that the range of colour temperatures capable of evoking a white sensation increases with increasing luminance.

For a single background stimulus, the neutral points lie close to the Munsell hue locus intercepting the background chromaticity. The blue display primary viewed alone produces an achromatic point which is outside the chromaticity gamut of the display. The blue display primary produces a greater neutral shift than the green primary which, in turn, is greater than that produced by the red phosphor. This finding is in agreement with the observations by Sugimoto et al (1973)[8] of the chromaticity of preferred human flesh colour. The inductive effect of a blue surround stimulus was significantly greater than for the other colours tested, with red backgrounds showing the smallest colour shifts. Kinney (1962)[35] also showed that stronger inductive effects were produced by blue and green surrounds than red or yellow surrounds of similar luminance and purity. In a further study, Kinney (1967)[30] showed that blue surround stimuli produced an immediate and almost complete adaptive effect, whereas that produced by red stimuli showed a strong time dependence.

Hunt (1975)[36] states that the reason for the choice of a television reference neutral matching daylight rather than incandescent light was that the surround luminance in daylight is usually much higher than that of the display, thereby dominating the viewers adaptation. Hence, a display 'white' point matching illuminant A would appear excessively yellow, whereas a display neutral matching illuminant D_{65} would not appear excessively blue in the low surround luminance conditions typical for domestic evening viewing. The measurements described in this chapter have shown that given a period for adaptation in conditions where the display scene is the dominant stimulus, a reference neutral point matching illuminant D_{65} or even of higher colour temperature is more likely to evoke a neutral sensation than one matching illuminant A.

An ambient illumination of 400 lux in the plane of the display screen produces a measurable shift in chromaticity due to light scattered from the display screen. This shift is in the direction of the illuminant chromaticity and therefore for daylight matching light sources produces a simple colour desaturation without hue shift. For conditions in which the display dominates the field of view, measured shifts in the achromatic point are consistent with a change of stimulus chromaticity rather than any peripheral adaptation.

The achromatic point is strongly dependent on luminance. As the luminance of a picture element is increased relative to the surround, the achromatic point appropriate to that element will shift towards the chromaticity of daylight, a characteristic of the aperture mode of appearance described in Appendix 5 . When the luminance of an element is reduced relative to that of the surround, the achromatic point moves closer to that of the dominant colour stimulus.

The position of the achromatic point produced by multiple colour stimuli is more difficult to predict. The use of the colour balance point, with adjustment for incomplete adaptation as suggested by Richards and Parks, was not always found to give accurate prediction of the achromatic point. However, for off-balance images, in which the colour balance point differs significantly from the reference neutral, the approximation is likely to produce a more accurate result than a simple assumption that the achromatic point remains coincident with the chromaticity of the reference neutral. The measured achromatic points show a bias towards the blue-green boundary of the display colour gamut.

When the colour balance point does coincide with the reference neutral chromaticity, the achromatic point will be close to this chromaticity. When a large area of the display has a chromaticity at or near the reference neutral point with average luminance, the position of the achromatic point, irrespective of the remaining colour stimuli, will remain relatively stable.

A two-stage model based on a non-linear three-receptor system is able to predict the achromatic stimulus for different levels of luminance. The model requires gain parameters to be defined for each colour balance point. The measured achromatic points behave in a manner which may help to promote constancy, as discussed in Appendix 5. As the luminance of a small, subjectively grey stimulus in the display is increased, keeping its chromaticity constant, the stimulus takes on the hue characteristic of the colour balance point viewed as an aperture stimulus. In many real scenes this colour balance point would be close to the chromaticity of the illuminant. Thus a small neutral surface simulated within the display may take on the hue of an apparent illuminant as its reflectance is increased, behaving in a similar manner to specular reflections, which are thought to promote constancy in real images by providing a cue to the colour of the illuminant.

The model presented in section 5.10 for colour prediction requires the achromatic point to be defined for the average luminance level within the display, subsequent colour predictions being based on the luminance relative to this level. A recent model, proposed by Nayatani et al, described in Appendix 5, is also based on a non-linear development of the von Kries coefficient law. However, by including luminance as an absolute quantity, the Nayatani model is able to predict changes of brightness and colour perception as illuminance is varied. It should therefore be possible using this model to predict the achromatic point directly for the display system with any level of luminance.

It is recommended that further study of colour prediction for the display system should include the Nayatani model.

References

1. Kelly, K. L.
J. Opt. Soc. Amer., vol. 33, p. 627 (1943).

2. Taylor, E. W.
Proc. of I.E.E., vol. 118, March/April, p. 449 (1971).

3. McMann, R. H. et al.
I.E.E. Conference Publication no. 119, International Broadcasting Convention (1974).

4. Bartleson, C. J. and Breneman, E. J.
Phot. Sci. Eng., vol. 11, p. 254 (1967).

5. Brown, G.
B.B.C. Research Report, 1971/4 (1971).

6. Bartleson, C. J. and Bray, C. P.
Phot. Sci. Eng., vol. 6, p. 19 (1962).

7. Bartleson, C. J.
Phot. Sci. Eng., vol. 10, p. 104 (1966).

8. Sugimoto, M. et al.
J. of S.M.P.T.E., vol. 82, p. 71 (1973).

9. Bartleson, C. J.
J. of S.M.P.T.E., vol. 77, p. 1 (1968).

10. Townsend, G. B.
I.E.E. Conference Report Series no. 5, Television Engineering, p. 528 (1962).

11. Carnt, P. S. and Townsend, G.B.
Subjective visual effects of A.N.T.S.C. Colour television receiver tolerances. Acta Electronica, vol. 2. p. 77 (1957).

12. Hurvich, L. M. and Jameson, D.
J. Opt. Soc. Am., vol. 41, 8, p. 528 (1951).

13. Winch, G. T.
G.E.C. Journal, vol. 16, p. 197 (1949).

14. MacAdam, D. L.
Proc. Inst. Radio Engnrs., vol. 43, p. 11 (1955).

15. Pearson, D. E.
Transmission and Display of Pictorial Information. Pentech Press Limited (1975).

16. Hunt, R. W. G.
The Reproduction of Colour. Fountain Press, p. 440 (1975).

17. Hunt, R.W.G.
see reference 16, p. 180.

18. Pearson, D.E.
see reference 15, p. 158.

19. Pearson, D.E. and Rubinstein, C.B.
J. of S.M.P.T.E., vol. 80, p. 15 (1971).

20. Judd, D.B.
J. of Opt. Soc. Am., vol. 39, p. 2 (1940).

21. Hunt, R.W.G. and Winter, L.M.
J. Photo. Sci., vol. 23, p. 112 (1975).

22. Richards, W. and Parks, E.A.
J. Opt. Soc. Am., vol. 61, 7, p. 971 (1971).

23. Pearson, D.E.
see reference 15, pp 144-156.

24. Wright, W.D.
Color Research and Application, vol. 6, 3(Fall, 1981).

25. Brown, W.R.J.
J. Opt. Soc. Am., vol. 42, p. 837 (1952).

26. Pearson, D.E.
see reference 15, p. 154.

27. MacAdam, D.L.
J. of S.M.P.T.E., vol. 57, p. 197 (1951).

28. Helson, H. and Michels, W.C.
J. Opt. Soc. Am., vol. 38, 12, p. 1025 (1948).

29. MacAdam, D.L.
J. Opt. Soc. Am., vol. 40, 9, p. 589 (1950).

30. Kinney, J.A.S.
Vision Res., vol. 7, p.299 (1967).

31. Hurvich, L.M. and Jameson, D.
J. Opt. Soc. Am., vol. 41, p. 787 (1957).

32. Land, E.H.
Proc. Nat. Acad. Sci., vol. 45, pp 115 and 636 (1959).

33. Pearson, D.E., Rubinstein, C.B. and Spivack, G.J.
J. Opt. Soc. Am., vol. 59, 5, p.644 (1969).

34. Judd, D.B.
J. Opt. Soc. Am., vol. 50, 3, p254 (1960).

35. Kinney, J.A.S.
Vision Res., vol. 2, p. 503 (1962).

36. Hunt, R.W.G.
see reference 16, p.439.

Chapter 6

CONCLUSIONS

6.1 Summary

A prototype display system has been described in this thesis which was developed as an aid to the teaching of lighting design in a Polytechnic Department of Architecture. The system produces an image which represents the distribution of chromaticity and relative luminance in a lighted room. The light distribution is calculated by a software inter-reflection model which takes account of both the spectral reflectance characteristics of the surfaces and the spatial and spectral distribution of the luminaire output. The colour rendering properties of the light source are therefore implicit in the calculation. By using hardware to define the image geometry, the system is able to operate inter-actively, the effect of design changes being seen without delay.

Objective measurements have shown the system to be capable of reproducing chromaticity and relative luminance to a high degree of accuracy. The system is not dependent on the computer hardware or specialised display system for generating the image; this gives four advantages;

(i) The cost of the system is kept low.

(ii) The system may be made portable.

(iii) The accuracy of the reproduced chromaticity and the spatial resolution are not subject to the constraints of a particular display system.

(iv) The image may be changed almost instantaneously, avoiding the 'wipe through' delay associated with general purpose high resolution graphics systems.

To summarise, the development of this system has demonstrated that with a suitable balance between hardware and software a solution to the display problem may be found which produces an image having the subjective resolution of a 64K byte frame store, which can be changed almost instantaneously, and which costs a fraction of the outlay necessary for a general purpose graphics system.

In broadcast television chromaticity is used as the basis for colour reproduction and for this reason chromaticity reproduction was used as the starting point in the design of the display system. The experience of broadcast colour television has shown, however, that accurate chromaticity reproduction may not be necessary and that some deliberate chromaticity distortion may produce a more pleasing image. However, more stringent requirements for perceived colour reproduction may apply in the case of the display system, which provides fewer perceptual cues than a real image. For this reason, in addition to objective colour measurement, subjective experiments were carried out to measure perceived colour.

Several workers in the field of perceived colour have highlighted the importance of the neutral or achromatic point to the perception of coloured stimuli, (e.g. Judd, Pearson, Hunt and Wright). By measuring the achromatic point, the relationship between displayed chromaticity and perceived colour has been determined using the display system with some specific images and viewing conditions. Using a small number of observers, shifts in the achromatic point were shown to be in qualitative agreement with those observed by MacAdam and Hunt and Winter for single colour stimuli. Luminance was found to be an important factor in the perception of neutral, as observed by Hunt and Winter. By applying a non-linear tri-receptor model as proposed by Richards and Parks, it was shown that the position of the achromatic point could be predicted as the luminance of the test area was varied relative to that of the background. These results support the proposition made by Richards and Parks that such a model, in addition to predicting the Land perceptions, could predict the perceived colour shifts produced by highly saturated stimuli.

The achromatic point has been measured using some multi-coloured images as well as for single colour stimuli. These measurements have shown that with the display as the only visual stimulus and where the average displayed chromaticity is some distance away from the daylight reference point, significant shifts of achromatic point could occur. Where a balanced colour combination was used, giving an average chromaticity close to the daylight point, the position of the achromatic point remained relatively stable, at or near this chromaticity.

The results of these perceived colour measurements have shown that for some images the reproduction in the display of the chromaticity in a real scene may not result in the same colour perception. By use of a predictive model it may be possible to compensate for these perceptual differences. Such a model requires further development to allow its general use as part of the display system.

6.2 Recommendations for future work.

It is probable that some simple additional image details may provide an increase in realism which outweighs the disadvantages of more complex hardware. For example, a doorway may give a useful reference to scale. In real rooms, wall luminance is usually higher near the ceiling than adjacent to the floor, showing a gradual fall between these extremes. These small luminance variations may contribute to the realism of the simulated image.

In the general case of a high resolution graphic display system these graduations, once calculated, would be displayed smoothly using the high density of available pixels. The luminance calculation would, however, involve increasing the number of elements in the inter-reflection model, resulting in an enormous increase in the volume of arithmetic and, unless the elements are made sufficiently small, a corresponding degrading of picture quality. A more economical approach might be to exploit the multiplying capability of the D/A converters to modulate the luminance. Since the luminance varies only slowly, the relatively low frequency response of the converter would not be a disadvantage. As this small luminance variation is used largely for cosmetic purposes, high levels of accuracy are unlikely to give any particular advantage. An approximation to the true function using an analogue function generator to drive the converter, may therefore provide an economical solution.

As already pointed out, the optimum system design involves identifying the minimum amount of picture detail to achieve the level of realism required for a particular application. It would be useful, therefore, to identify formally the minimum requirements for each application, using model rooms and closed circuit television images of real rooms as references.

Given the limited number of visual cues from the graphic image, the accuracy with which surface lightness can be perceived, and how this is influenced by additional image detail, should be investigated.

In judging a lighting design, the modelling of an object within the lighted space is usually an important consideration. A useful addition to the display system might, therefore, be a simulated object, such as a sphere, polyhedron or even a human face, to give the viewer more cues to the direction of the flow of light. Such an image could be generated by storing, using a PROM, the appearance of a face at two extremes of vector/scalar ratio, intermediate values being generated by video mixing.

As emphasised in the conclusions, the position of the achromatic point is central to the mechanism of colour perception. The model investigated in Chapter 5 showed that for simple images it is possible to predict how a colour stimulus will be perceived, even when the achromatic point is displaced some distance from the daylight point. This preliminary work suggests that such a model could be extended to cover the general case of a simulated room image. The influence of ambient lighting outside the display on the stability of the achromatic point also requires further investigation.

To date, no experimental work has been attempted to validate the system formally, either by using models or real scenes viewed via closed circuit television. The system has, however, been used by both students or architecture and lighting engineers, from which experience its potential has been clearly demonstrated.

DECLARATION

No part of this thesis has been submitted in support of an application for another award of the C.N.A.A. or qualification at any university or other institution of learning.

D. J. Gilderdale

ACKNOWLEDGEMENTS

I would like to thank Dr. E. MacQuade and Mr. J. Lynes for their supervision and guidance during the course of this work and the preparation of this thesis.

I am indebted to the School of Electrical and Electronic Engineering, Plymouth Polytechnic for the use of its facilities, to Devon County Council for its financial support and to my colleagues who contributed in various ways to the completion of this work.

Particular thanks are due to Jacqueline Pennock and Ann Freemantle for their patience and encouragement during the preparation of the typescript.

ADVANCED STUDIES undertaken during the programme of research.

Short course in Educational Technology,
Plymouth Polytechnic, November 1974.

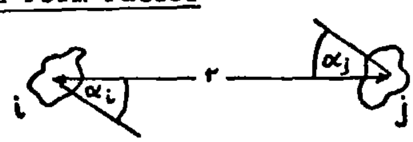
Short course in Microprocessors,
Plymouth Polytechnic, April 1977.

Advanced lighting calculations, Illuminating Engineering Society Course,
University of Warwick, March 1977.

Colour technology course,
Instrumental Colour Systems Limited, November 1977.

I.E.E. Conference on visual displays,
University of Lancaster, 1978.

Exchange Function and Form Factor



Consider 2 infinitesimal elements of two interacting planar surfaces as shown above. Let surface j have a luminance L_j. Then the flux F_{ij} which is intercepted by element i will be given by the inverse square law:-

$$F_{ij} = \frac{L_j \cos \alpha i \cos \alpha j}{r^2}$$

where the surfaces are assumed to be uniform diffusing Lambertian surfaces. (A Lambertian surface is one which diffuses the light uniformly, distributing it about the normal to the surface as the cosine of the angle of the reflected ray from the normal, regardless of the angle of incidence.) Practical surfaces reflect neither 100% diffused light nor 100% specular light, but a combination of the two. However matt paint surfaces approximate quite well to perfect diffusers and all surfaces considered here are assumed to be such.

The total flux intercepted by surface i from surface j is thus given by

$$F_{ij} = \int_{A_j} \int_{A_i} \frac{L_j \cos \alpha i \cos \alpha j}{r^2} \, dA_i dA_j$$

For conciseness Dourgnon (1955) introduces a function a_{ij} called the exchange function where

$$a_{ij} = \frac{1}{\pi} \int_{A_j} \int_{A_i} \frac{\cos \alpha i \cos \alpha j}{r^2} \, dA_i dA_j$$

(the factor $\frac{1}{\pi}$ is introduced here in order to simplify the units of luminance when L_j is expressed in cd/unit area.)

The fraction of the total flux F_j which intercepts surface i is therefore

$$\frac{F_{ij}}{F_j} = f_{ji} = \frac{aij}{A_j}$$ ($F_j = L_j \cdot A_j$ where L_j is expressed in lumens/unit area).

the function f_{ji} is called a form factor and may thus be defined as

$$f_{ji} = \frac{1}{\pi A_j} \int_{A_j} \int_{A_i} \frac{\cos \alpha i \cos \alpha j}{r^2} dA_i dA_j$$

APPENDIX 2B

The Interreflectance method of lighting calculation

This approach to solution of the lighting problem, (Moon and Spencer 1946) is based on direct solution of the Fredholm integral equation and whilst avoiding some errors implicit in the finite difference technique still requires that certain approximations be made which limit the final accuracy.

The case of a rectangular cylinder

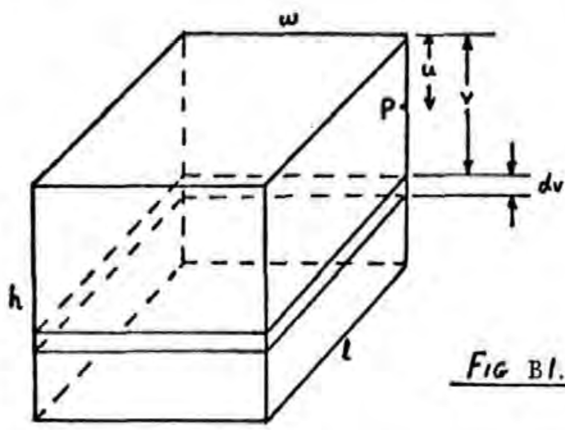


FIG B1.

A right cylinder of rectangular section is shown in Fig. B1.

The illuminance at P may be written as

$$E(s) = E_o(s) + \int_o^{h/w} K\,(s,t)\;E\,(t)dt \qquad (s = u/w \quad t = v/w) \qquad (1)$$

the function K(s,t) is related to the amount of luminous flux emanating from the element which intercepts P. The derivation of K(s,t) (Parry Moon 1940) results in an expression of the form.

$$K(s,t) = \frac{1}{\pi(1+1/_w)}\left\{ \ln\left[\frac{(1/w)^2 + (s-t)^2}{(s-t)^2}\right] - \ln\left[\frac{1 + (1/w)^2 + (s-t)^2}{1 + (s-t)^2}\right] + (1/w)\frac{1}{\left[1 + (s-t)^2\right]^{3/2}} \tan^{-1}\frac{(1/w)}{\left[1 + (s-t)^2\right]^{\frac{1}{2}}} + \frac{(1/w)}{\left[(1/w)^2 + (s-t)^2\right]^{3/2}} \tan^{-1}\frac{1}{\left[(1/w)^2 + (s-t)^2\right]^{\frac{1}{2}}} \right\} \qquad (2)$$

When this expression for K(s,t) is substituted into equation (1) the resulting integral does not have a rigorous solution which may be obtained directly. (Parry Moon 1946).

However, Buckley (1927) showed that for the case of a circular cylinder K(s,t) may be approximated by an exponential function. This technique is applied to cylindrical shapes of other cross-sections by Moon and Spencer (1946). The expression for K(s,t) in equation (2) is approximated by:

$$K(s,t) = Ae^{-a(s-t)} \quad \text{where A and a are constants.}$$

In practice $a = \frac{pw}{2S}$ $A = a/_2$

where p = perimeter of a horizontal cross-section

S = area of the cross-section

W = width of cylinder

When this expression for K(s,t) is substituted in (1) a solution is obtained of form

$$\frac{E(s)}{E_o(s)} = \frac{1}{1-\rho} + B\cosh(k\ s) + C\sinh(k\ s) \quad (k=a(1-\rho)^{\frac{1}{2}})$$

The accuracy of this result is obviously a function of the accuracy of the approximation for K(s,t).

Fig. B2 gives a graphical comparison of the true function K(s,t) and its exponential approximation.

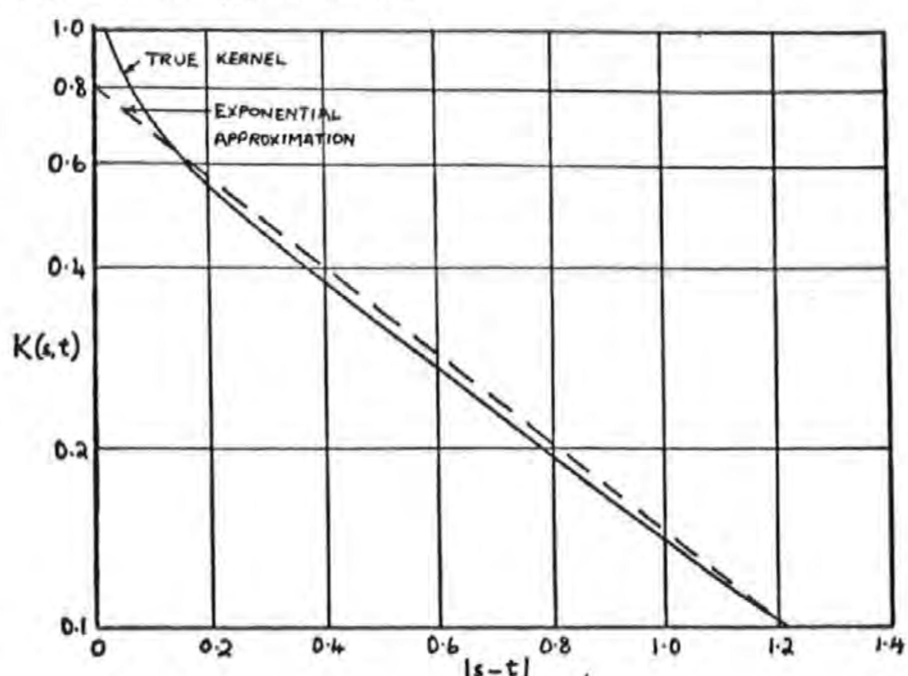


FIG B2. KERNELS FOR A RECTANGULAR CYLINDER FOR $w/l = 2.0$

APPENDIX 2C

Calculation of the total flux passing from one surface element to another

The theory of interreflections depends on a knowledge of the fraction of the total flux leaving one surface element which is incident on another. This fraction we call the "form factor" for the pair of surface elements.

The general planar case shown in Fig. 2.0 shows two surface elements of random orientation.

The fraction of flux from surface n which intercepts surface i is given by:

$$fin = \frac{1}{A_i} \int_{A_i} \int_{A_n} \frac{\cos \theta_i \cos \theta_n}{\pi r^2} dA_i dA_n$$

This integral has been calculated for various geometries (Hamilton & Morgan 1952, Philips & Prokhovnik 1960, Coomber & Jay 1967.)

For the case of a rectangular room, only two geometries are of interest as shown in Fig. C1.

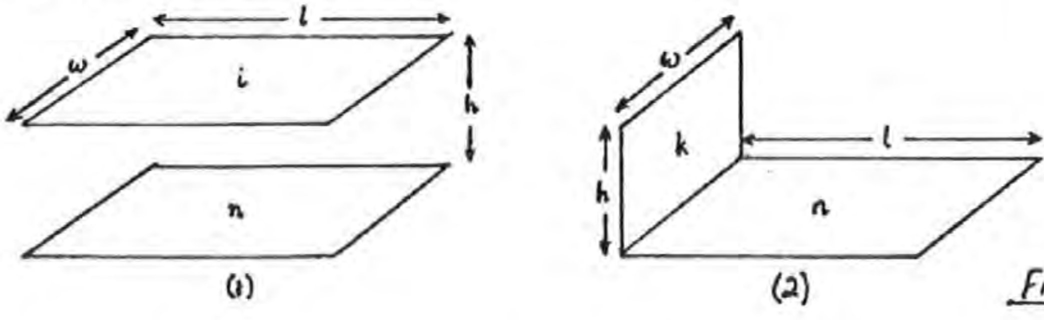


FIG C1.

The form factor for case (1), surface i parallel to surface n

$$fin = \frac{2}{\pi}\left\{\frac{(h^2+l^2)^{\frac{1}{2}}}{l} \tan^{-1}\frac{w}{(h^2+l^2)^{\frac{1}{2}}} + \frac{(h^2+w^2)^{\frac{1}{2}}}{w} \tan^{-1}\frac{l}{(h^2+w^2)^{\frac{1}{2}}} - \frac{h}{l}\tan^{-1}\frac{w}{h} - \frac{h}{w}\tan^{-1}\frac{l}{h} - \frac{1}{2}\frac{h^2}{lw}\ln\frac{h^2(h^2+l^2+w^2)}{(h^2+w^2)(h^2+l^2)}\right\} \qquad (1)$$

For case (2), surface k perpendicular to surface n

$$f_{kn} = \frac{1}{\pi}\left\{\tan^{-1}\frac{1}{h} + \frac{w}{h}\tan^{-1}\frac{1}{w} - \frac{(h^2+w^2)^{\frac{1}{2}}}{h}\tan^{-1}\frac{1}{(h^2+w^2)^{\frac{1}{2}}} + \frac{h}{4l}.\ln\frac{(h^2(h^2+l^2+w^2))}{(h^2+w^2)(h^2+l^2)} - \frac{1}{4h}.\ln\frac{l^2(h^2+l^2+w^2)}{(h^2+l^2)(l^2+w^2)} + \frac{w^2}{4lh}.\ln\frac{w^2(h^2+l^2+w^2)}{(h^2+w^2)(l^2+w^2)}\right\} \quad (2)$$

Hence, for a given room geometry the form factors may be derived by simple substitution. It is perhaps significant to note that the form factors are functions of the proportions of the rectangular enclosure, rather than the absolute sizes.

For a rectangular parallelopiped there are 36 form factors f_{ij} required to discribe the complete interaction of flux. However, certain other relationships are valid viz:

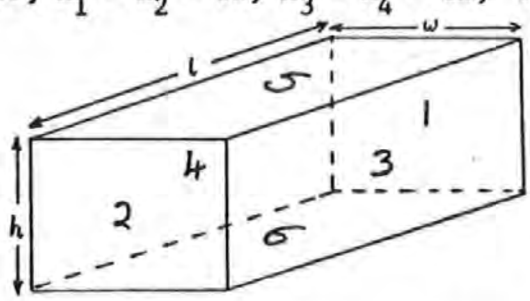
(1) $f_{ii} = 0$ (a plane cannot illuminate itself)

(2) $A_i f_{ij} = A_j f_{ji}$ (reciprocity relation)

(3) $\sum_{k=1}^{6} fik = 1$ (summing total flux)

(4) $f_{13} = f_{14},\ f_{15} = f_{16},\ f_{35} = f_{36}$
$f_{23} = f_{24},\ f_{25} = f_{26},\ f_{45} = f_{46}$ } Symmetry

also, of course, $A_1 = A_2 = wh$; $A_3 = A_4 = lh$; $A_5 = A_6 = lw$.



Therefore, for the 36 form factors it is possible to find 33 interrelations from the above equations. This leaves only three of the factors which are independent and having calculated these three using formulee (1) and (2) above, the others may be quickly derived using relationships (3) and (4).

APPENDIX 3A

Setting-up procedure for the display system.

The following is a summary of the procedures required for set-up and calibration of the display system. If complete specification of the display tube characteristics is available, steps 12 and 13 may be omitted.

1. The convergence and purity of the CRT display are adjusted using the procedures described in the maintenance manual. A television pattern generator is required for this procedure.

2. Set the pattern generator to produce a grey scale. The generator output should be applied to the red, green and blue channels connected in parallel.

3. Set the display contrast control to zero and adjust the brightness control to the zero luminance threshold using a photometer.

4. Increase the contrast setting to produce a full grey scale with peak luminance ≃ 70 cd/m².

5. Using a colorimeter, adjust the chromaticity of full white, using the chroma gain controls R,G,B to produce a match for D_{65}.

6. Adjust the display bias controls R, G and B to obtain D_{65} chromaticity for low luminance near the bottom of the grey scale.

7. Repeat steps 5 and 6 until each luminance step matches D_{65}.

8. Having achieved grey scale tracking for the display monitor, connect the colour display generator, with the monitor R,G,B inputs set to high impedance.

9. Set the display monitor contrast control to minimum and use the software test routine to set the input of each D/A converter to 255_{10}.

10. Using a digital voltmeter, set the output of each D/A buffer amplifier in turn to 1.00 volts, a total of fifteen adjustments.

11. Increase the display 'contrast' control setting, if all is well peak white should be displayed over the whole screen, with the chromaticity matching D_{65}. The 'contrast' control should be again adjusted to give a peak white luminance of 75 cd/m².

12. Measure the chromaticities of the display CRT phosphors using the methods discussed in section 3.33.

13. The non-linearity of the display tube is measured using the procedure described in section 3.32, 'Gamma correction'.

The final two measurements, steps 12 and 13, may be omitted if detailed manufacturer's data is available defining the CRT phosphor chromaticities and gamma characteristics.

APPENDIX 3B

Sources of error in chromaticity measurement.

Measurement from television cathode ray tubes adds its own peculiar problems to the sources of error normally encountered in colour measurement. Some of these problems are outlined in the following sections along with precautions which help to maintain accuracy.

1. Photomultiplier dark current.

With no light input a photomultiplier tube passes a small residual current which usually varies with temperature. This dark current adds an error term to all the readings, which, if not taken into account, will reduce the accuracy of the computed chromaticity co-ordinates, giving an artificially desaturated result. Dark current is simply eliminated by subtraction from $R(\lambda)$ for each wavelength before computing chromaticity. Frequent monitoring is necessary to avoid the effect of drift resulting from temperature changes.

2. Drifts in display screen luminance.

The screen luminance is dependent on:

(i) The beam current of the CRT, controlled by the grid and first-anode voltage

(ii) The E.H.T. voltage

(iii) The scan repetition rate and spot writing speed.

Temperature changes and small changes in supply voltage can have a significant effect on the above three parameters. The complete solution to this problem is provided by a feedback loop in which the screen luminance is constantly monitored by a photocell, the output of which is used to control the beam current.

3. Hum bars on the display.

A variation in the d.c. supply voltages used for the display circuitry, due, for example, to inadequate smoothing, or coupling with mains frequency circuitry, can result in a periodic change of display luminance having the appearance of horizontal bars which move up or down the screen. For the purposes of chromaticity measurement, this problem may be eliminated by synchronizing the frame repetition frequency to that of the mains supply.

4. Purity errors in 3-colour CRTs.

If the cathode ray tube electron beams are incorrectly aligned, it is possible for the wrong phosphor dots to be excited, resulting in light output of colours other than the intended primary. Before any photometric measurements are made it is therefore necessary to ensure that the purity controls of the display monitor are correctly adjusted. Also, the scan amplitude of the display is best kept within the screen area since electrons striking the walls of the tube may be scattered, causing random excitation of phosphor dots and hence colour desaturation.

5. The effect of stray light leakage.

Provided that adequate precautions are taken to shield the display tube and the space between it and the monochromator from external light, it is unlikely that stray light will affect the measurements. The effects of stray light can be distinguished from other errors such as photomultiplier dark current by the variation in response as the monochromator wavelength setting is varied with zero display screen luminance.

6. Second order spectral responses.

The equation describing the relationship between wavelength and angle of incidence for a diffraction grating takes the form:

$$a(\sin \theta - \sin \theta_i) = n\lambda \qquad (\theta_i \text{ is the angle of incidence})$$

The output spectrum is repeated for integer values of n. For a given response at angle θ, a similar, although attenuated, output will also be present at 2θ, 3θ etc. For a given wavelength λ, the output could therefore be the result of radiation at λ, $\lambda/2$, $\lambda/3$ etc. In the general case of spectral analysis over a wide range of wavelength, anomalous results may be produced by these higher order responses. Higher order spectra may be identified by the use of bandpass filters at submultiples of the frequency of interest. In the practical measurement of the chromaticity of television cathode ray tube phosphors, the light output outside the range 380-780 nm is negligible and errors due to second order spectra can be neglected.

The above sections summarise the potential sources of error encountered when measuring chromaticity using wavelength by wavelength spectral analysis. The calibration of the monochromator was discussed earlier. Stability of the voltmeter and monochromator are both higher than other aspects of the measurement. Absolute calibration of the output voltage is not necessary as only the ratios of readings are required for chromaticity measurement. The most important source of error is the change of screen luminance resulting from both supply voltage and temperature variations, which over the period of twenty minutes needed for the complete measurement can be quite large. Since the display monitor was available only on loan it was not possible to arrange for a feedback loop to regulate the beam current automatically. However, care was taken to allow a warm-up period of one hour in a dark-room having a near constant ambient temperature. The whole system was operated via a regulated mains supply.

APPENDIX 3C

COMPUTER PROGRAMS

Computer program for the general case of inter-reflection taking account of illuminant and surface colour.

```
DX1:MBIN1.BDJ
10 DIM P(6),G(6),S(6),O1(6),O2(6),O3(6),A(6,7),B(3),X(6),Y(6),Z(6)
12 PRINT CHR$(31);
14 PRINT CHR$(29); \ FOR I=1 TO 21 \ R1=0 \ GOSUB 800 \ NEXT I
16 R1=25 \ GOSUB 800 \ R1=15 \ GOSUB 800 \ R1=15 \ GOSUB 800 \ R1=25 \ GOSUB 800 \ R1=25 \ GOSUB 800 \ R1=120 \ GOSUB 800
18 PRINT CHR$(31);
50 PRINT  \ PRINT "ROOM DIMENSIONS (IN METRES)?"
52 PRINT  \ PRINT "LENGTH ="; \ INPUT #0,L
54 PRINT  \ PRINT "WIDTH ="; \ INPUT #0,W
56 PRINT  \ PRINT "HEIGHT ="; \ INPUT #0,H
60 A1=L*W \ K1=A1/(H*(L+W)) \ GOSUB 500
65 PRINT  \ PRINT "NOW THE LIGHTING SYSTEM"
67 PRINT  \ PRINT "BZ NO? ="; \ INPUT #0,B1 \ GOSUB 700
69 PRINT  \ PRINT "FLUX FRACTION RATIO ="; \ INPUT #0,F
71 PRINT  \ PRINT "LUMINAIRE OUTPUT IN WATTS ="; \ INPUT #0,O
73 PRINT  \ PRINT "NO OF LUMINAIRES ="; \ INPUT #0,Z \ L1=O*80*Z/(1+F)
75 FOR I=1 TO 6 \ X(I)=0 \ Y(I)=0 \ Z(I)=0 \ NEXT I
80 N=6 \ OPEN "DX1:RSUR1.BDJ" AS FILE VF1(269)
82 OPEN "DX1:DL6.BDJ" AS FILE VF2(44)
85 OPEN "DX1:RLD1.BDJ" AS FILE VF3(269)
87 FOR C=0 TO 44 \ P=VF2(C) \ FOR I=0 TO 5 \ P(I+1)=VF1(I+C*6) \ NEXT I
89 FOR J=1 TO 2
91 A(J,3-J)=-G(1)
93 A(J+2,5-J)=-G(2)
95 A(J+4,7-J)=-G(3) \ NEXT J
100 FOR I=0 TO 1 \ FOR J=0 TO 1
102 A(I+3,2-J)=-B(1)*G(4)
104 A(I+5,2-J)=-B(2)*G(5)
106 A(I+1,4-J)=-G(4)
108 A(I+5,4-J)=-B(3)*G(6)
110 A(I+1,6-J)=-G(5)
111 A(I+3,6-J)=-G(6) \ NEXT J \ NEXT I
120 FOR I=1 TO 6 \ A(I,I)=1/P(I) \ NEXT I
122 FOR I=1 TO 4 \ A(I,7)=L1*P*(1-D)*K1/(2*A1) \ NEXT I
124 A(5,7)=L1*P*F/A1
126 A(6,7)=L1*P*D/A1
130 GOSUB 310 \ FOR J=0 TO 5 \ VF3(J+C*6)=S(J+1) \ NEXT J
132 NEXT C
140 OPEN "DX1:CMX.BDJ" AS FILE VF4(44) \ OPEN "DX1:CMY.BDJ" AS FILE VF5(44)
142 OPEN "DX1:CMZ.BDJ" AS FILE VF6(44)
144 FOR I=0 TO 44 \ FOR C=0 TO 5 \ X(C+1)=X(C+1)+(VF3(I*6+C)*VF4(I))
146 Y(C+1)=Y(C+1)+(VF3(I*6+C)*VF5(I)) \ Z(C+1)=Z(C+1)+(VF3(I*6+C)*VF6(I)) \ NEXT C \ NEXT I
150 FOR I=1 TO 6
151 X(I)=X(I)*3.38306 \ Y(I)=Y(I)*3.38306 \ Z(I)=Z(I)*3.38306
152 O1(I)=3.90445*X(I)-2.04704*Y(I)-.608984*Z(I)
```

```
154 O2(I)=-1.13404*X(I)+2.00904*Y(I)+.0629243*Z(I)
156 O3(I)=.063369*X(I)-.20609*Y(I)+1.05279*Z(I)
158 NEXT I
160 PRINT CHR$(29);
162 R1=O1(1)^.47619 \ GOSUB 800 \ FOR I=3 TO 6 \ R1=O1(I)^.47619 \ GOSUB 800 \ NEXT I
164 R1=O2(1)^.47619 \ GOSUB 800 \ FOR I=3 TO 6 \ R1=O2(I)^.47619 \ GOSUB 800 \ NEXT I
166 R1=O3(1)^.47619 \ GOSUB 800 \ FOR I=3 TO 6 \ R1=O3(I)^.47619 \ GOSUB 800 \ NEXT I
168 FOR I=1 TO 3 \ R1=0 \ GOSUB 800 \ NEXT I
170 PRINT CHR$(31);
192 CLOSE
194 END
310 FOR K=1 TO N-1 \ FOR J=K+1 TO N \ GOSUB 340 \ NEXT J \ NEXT K
315 S(N)=A(N,N+1)/A(N,N)
320 FOR K=1 TO N-1 \ GOSUB 345 \ NEXT K \ RETURN
340 M=A(J,K)/A(K,K) \ FOR J1=K TO N+1 \ A(J,J1)=A(J,J1)-A(K,J1)*M \ NEXT J1 \ RETURN
345 M=A(N-K,N+1) \ FOR J=0 TO K-1 \ M=M-A(N-K,N-J)*S(N-J) \ NEXT J
350 S(N-K)=M/A(N-K,N-K) \ RETURN
500 I=1 \ X=H/L \ Y=W/L \ B(1)=Y \ B(2)=X \ GOSUB 550
504 X=H/W \ Y=L/W \ B(3)=X \ I=I+1 \ GOSUB 550
508 X=L/W \ Y=W/H \ I=I+1
550 A=LOG(SQR((1+X^2)*(1+Y^2)/(1+X^2+Y^2)))
552 A=A+Y*SQR(1+X^2)*ATN(Y/SQR(1+X^2))
554 A=A+X*SQR(1+Y^2)*ATN(X/SQR(1+Y^2))
556 A=A-Y*ATN(Y)-X*ATN(X)
558 A=A*.6366/(X*Y)
560 IF I=1 GO TO 570
562 IF I=2 GO TO 572
564 IF I=3 GO TO 574
570 G(1)=A \ A(1,4)=(1-A)/2 \ RETURN
572 G(2)=A \ A(2,4)=(1-A)/2 \ RETURN
574 G(3)=A \ A(3,4)=(1-A)/2
580 A(1,1)=1 \ A(3,1)=0 \ A(1,2)=1 \ A(2,2)=0 \ A(1,3)=0 \ A(2,3)=1
582 A(2,1)=B(1) \ A(3,2)=B(2) \ A(3,3)=B(3)
584 N=3 \ GOSUB 310
586 FOR J=1 TO 3 \ G(J+3)=S(J) \ NEXT J \ RETURN
700 IF B1=1 THEN 704 \ IF B1=2 THEN 706 \ IF B1=3 THEN 708 \ IF B1=4 THEN 710 \ IF B1=5 THEN 712
702 IF B1=6 THEN 714 \ IF B1=7 THEN 716 \ IF B1=8 THEN 718 \ IF B1=9 THEN 720 \ GO TO 722
704 D=.093096+1.1171*K1-.70802*K1^2+.23617*K1^3-.039112*K1^4+2.52120E-03*K1^5 \ RETURN
706 D=7.81130E-04+1.1483*K1-.70015*K1^2+.22823*K1^3-.037249*K1^4+2.37800E-03*K1^5 \ RETURN
708 D=.026025+.95007*K1-.51692*K1^2+.15527*K1^3-.023885*K1^4+1.46000E-03*K1^5 \ RETURN
710 D=-.027528+.92475*K1-.48374*K1^2+.14254*K1^3-.021733*K1^4+1.32410E-03*K1^5 \ RETURN
712 D=8.10370E-04+.77028*K1-.36218*K1^2+.098212*K1^3-.013993*K1^4+8.06260E-04*K1^5 \ RETURN
714 D=-.026386+.63699*K1-.27732*K1^2+.073227*K1^3-.010398*K1^4+6.03620E-04*K1^5 \ RETURN
716 D=-.037394+.54261*K1-.21378*K1^2+.053176*K1^3-7.29080E-03*K1^4+4.14580E-04*K1^5 \ RETURN
```

```
718 D=-.051811+.49576*K1-.18267*K1^2+.043435*K1-5.79100E-03*K1^4+3.24420E-04*K1^5 \ RETURN
720 D=-.12453+.57149*K1-.25333*K1^2+.073761*K1^3-.011629*K1^4+7.37490E-04*K1^5 \ RETURN
722 D=-.1149+.39415*K1-.093688*K1^2+.01081*K1^3-1.61790E-04*K1^4-4.60100E-05*K1^5 \ RETURN
800 R4=INT(R1/32) \ R2=R1+96-(32*R4) \ R4=R4+32
802 PRINT CHR$(R4); \ PRINT CHR$(R2); \ RETURN
```

Computer program using full inter-reflection theory with a three-band colour approximation.

```
DX1:FIR2.BDJ
5 DIM X(6),Y(6),C(2),P(26),T(26),S(6),A(6,7),R(6),F(6),B(3),G(6)
6 DIM I(26),D(26)
7 PRINT CHR$(31);
9 PRINT CHR$(29); \ FOR I=1 TO 21 \ R1=0 \ GOSUB 300 \ NEXT I
11 F(1)=25 \ F(2)=15 \ F(3)=15 \ F(4)=25 \ F(5)=25 \ F(6)=120
14 PRINT CHR$(31)
16 PRINT  \ PRINT "ROOM DIMENSIONS (IN METRES)?"
18 PRINT  \ PRINT "LENGTH ="; \ INPUT #0,L
20 PRINT  \ PRINT "WIDTH ="; \ INPUT #0,W
22 PRINT  \ PRINT "HEIGHT ="; \ INPUT #0,H
24 A1=L*W \ K1=A1/(H*(L+W)) \ GOSUB 500
26 IF S$="DIM" THEN 60
28 FOR I=1 TO 6
30 GOSUB 260
32 PRINT  \ PRINT "REFLECTANCE= "; \ INPUT #0,R(I)
34 PRINT  \ PRINT "CHROMATICITY X= "; \ INPUT #0,X(I)
36 PRINT TAB(13);"Y= "; \ INPUT #0,Y(I)
38 C(1)=X(I) \ C(2)=Y(I) \ GOSUB 280
40 NEXT I \ PRINT  \ PRINT
42 PRINT "NOW THE LIGHTING SYSTEM"
44 GOSUB 130
46 PRINT  \ PRINT "LUMINAIRE OUTPUT IN WATTS= ";
48 INPUT #0,O \ PRINT  \ PRINT "NO. OF LUMINAIRES= ";
50 INPUT #0,Z
54 PRINT  \ PRINT "FLUX FRACTION RATIO= "; \ INPUT #0,F
56 T2=0
58 FOR K2=0 TO 2
60 FOR J=1 TO 2
62 A(J,3-J)=-G(1)
64 A(J+2,5-J)=-G(2)
66 A(J+4,7-J)=-G(3) \ NEXT J
70 FOR I=0 TO 1 \ FOR J=0 TO 1
72 A(I+3,2-J)=-B(1)*G(4)
74 A(I+5,2-J)=-B(2)*G(5)
76 A(I+1,4-J)=-G(4)
78 A(I+5,4-J)=-B(3)*G(6)
80 A(I+1,6-J)=-G(5)
82 A(I+3,6-J)=-G(6) \ NEXT J \ NEXT I
85 L1=O*80*Z/(1+F)
90 FOR J=1 TO 4 \ A(J,7)=L1*(1-D)*K1/(2*A1) \ NEXT J
92 A(5,7)=L1*F/A1
94 A(6,7)=L1*D/A1
```

```
98 FOR J=1 TO 6 \ A(J,J)=1/P(J+K2*10) \ NEXT J
100 N=6 \ GOSUB 310 \ FOR J=1 TO 6 \ T(J+K2*10)=S(J) \ NEXT J \ NEXT K2
102 GO TO 1000
104 PRINT  \ PRINT "C,LO,LS,FFR OR BZ? "; \ INPUT #0,S$ \ T2=-1
106 IF S$="C" THEN 120
108 IF S$="LO" THEN 124
110 IF S$="LS" THEN 126
112 IF S$="FFR" THEN 128
114 IF S$="BZ" THEN  GOSUB 130
115 IF S$="DIM" THEN 16
116 IF S$="SEP" THEN 2000
118 GO TO 58
120 PRINT  \ PRINT "SURFACE 1,2,3,4,5 OR 6? "; \ INPUT #0,I
121 PRINT  \ PRINT  \ PRINT "R AND X,Y FOR THE" \ GOSUB 260
122 PRINT "R= "; \ INPUT #0,R(I) \ PRINT TAB(16);"X= "; \ INPUT #0,X(I) \ PRINT TAB(16);"Y= "; \ INPUT #0,Y(I)
123 C(1)=X(I) \ C(2)=Y(I) \ GOSUB 280 \ GO TO 116
124 PRINT  \ PRINT "NEW LUMINAIRE OUTPUT IN WATTS= "; \ INPUT #0,O
125 PRINT  \ PRINT "NO. OF LUMINAIRES? "; \ INPUT #0,Z \ GO TO 58
128 PRINT  \ PRINT  \ PRINT "NEW FLUX FRACTION RATIO="; \ INPUT #0,F \ GO TO 116
130 PRINT  \ PRINT  \ PRINT "BZ NO?= "; \ INPUT #0,B1
132 IF B1=1 THEN 136 \ IF B1=2 THEN 138 \ IF B1=3 THEN 140 \ IF B1=4 THEN 142 \ IF B1=5 THEN 144
134 IF B1=6 THEN 146 \ IF B1=7 THEN 148 \ IF B1=8 THEN 150 \ IF B1=9 THEN 152 \ GO TO 154
136 D=.093096+1.1171*K1-.70802*K1^2+.23617*K1^3-.039112*K1^4+2.52120E-03*K1^5 \ RETURN
138 D=7.81130E-04+1.1483*K1-.70015*K1^2+.22823*K1^3-.037249*K1^4+2.37800E-03*K1^5 \ RETURN
140 D=.026025+.95007*K1-.51692*K1^2+.15527*K1^3-.023885*K1^4+1.46000E-03*K1^5 \ RETURN
142 D=-.027528+.92475*K1-.48374*K1^2+.14254*K1^3-.021733*K1^4+1.32410E-03*K1^5 \ RETURN
144 D=8.10370E-04+.77028*K1-.36218*K1^2+.098212*K1^3-.013993*K1^4+8.06260E-04*K1^5 \ RETURN
146 D=-.026386+.63699*K1-.27732*K1^2+.073227*K1^3-.010398*K1^4+6.03620E-04*K1^5 \ RETURN
148 D=-.037394+.54261*K1-.21378*K1^2+.053176*K1^3-7.29080E-03*K1^4+4.14580E-04*K1^5 \ RETURN
150 D=-.051811+.49576*K1-.18267*K1^2+.043435*K1^3-5.79100E-03*K1^4+3.24420E-04*K1^5 \ RETURN
152 D=-.12453+.57149*K1-.25333*K1^2+.073761*K1^3-.011629*K1^4+7.37490E-04*K1^5 \ RETURN
154 D=-.1149+.39415*K1-.093688*K1^2+.01081*K1^3-1.61790E-04*K1^4-4.60100E-05*K1^5 \ RETURN
260 IF I=1 THEN 262 \ IF I=2 THEN 264 \ IF I=3 THEN 266 \ IF I=4 THEN 268 \ IF I=5 THEN 270 \ GO TO 272
262 PRINT  \ PRINT "FACING WALL (1) "; \ RETURN
264 PRINT  \ PRINT "WALL BEHIND YOU (2) "; \ RETURN
266 PRINT  \ PRINT "RIGHT HAND WALL (3) "; \ RETURN
268 PRINT  \ PRINT "LEFT HAND WALL (4) "; \ RETURN
270 PRINT  \ PRINT "CEILING (5) "; \ RETURN
272 PRINT  \ PRINT "FLOOR (6) "; \ RETURN
280 A(1,1)=1 \ A(1,2)=1 \ A(1,3)=1
282 A(2,1)=1.7765-C(1)/.349 \ A(2,2)=.5497-C(1)/.584
284 A(2,3)=2.4603-C(1)/.063
286 A(3,1)=1-C(2)/.349 \ A(3,2)=1-C(2)/.584 \ A(3,3)=1-C(2)/.063
288 A(1,4)=R(I) \ A(2,4)=0 \ A(3,4)=0
290 N=3 \ GOSUB 310 \ P(I)=S(1)*4.83513 \ P(I+10)=S(2)*1.39594 \ P(I+20)=S(3)*13.0178 \ RETURN
```

```
300 R4=INT(R1/32) \ R2=R1+96-(32*R4) \ R4=R4+32
302 PRINT CHR$(R4); \ PRINT CHR$(R2); \ RETURN
310 FOR K=1 TO N-1 \ FOR J=K+1 TO N \ GOSUB 340 \ NEXT J \ NEXT K
315 S(N)=A(N,N+1)/A(N,N)
320 FOR K=1 TO N-1 \ GOSUB 345 \ NEXT K \ RETURN
340 M=A(J,K)/A(K,K) \ FOR J1=K TO N+1 \ A(J,J1)=A(J,J1)-A(K,J1)*M \ NEXT J1 \ RETURN
345 M=A(N-K,N+1) \ FOR J=0 TO K-1 \ M=M-A(N-K,N-J)*S(N-J) \ NEXT J
350 S(N-K)=M/A(N-K,N-K) \ RETURN
410 F(1)=24.08 \ F(2)=16.0533 \ F(3)=8.02667 \ F(4)=12.04 \ F(5)=51.1875 \ F(6)=87.75 \ GO TO 1000
412 F(1)=12.04 \ F(2)=8.02667 \ F(3)=16.0533 \ F(4)=24.08 \ F(5)=51.1875 \ F(6)=87.75 \ GO TO 1000
500 I=1 \ X=H/L \ Y=W/L \ B(1)=Y \ B(2)=X \ GOSUB 550
504 X=H/W \ Y=L/W \ B(3)=X \ I=I+1 \ GOSUB 550
508 X=L/W \ Y=W/H \ I=I+1
550 A=LOG(SQR((1+X^2)*(1+Y^2)/(1+X^2+Y^2)))
552 A=A+Y*SQR(1+X^2)*ATN(Y/SQR(1+X^2))
554 A=A+X*SQR(1+Y^2)*ATN(X/SQR(1+Y^2))
556 A=A-Y*ATN(Y)-X*ATN(X)
558 A=A*.6366/(X*Y)
560 IF I=1 GO TO 570
562 IF I=2 GO TO 572
564 IF I=3 GO TO 574
570 G(1)=A \ A(1,4)=(1-A)/2 \ RETURN
572 G(2)=A \ A(2,4)=(1-A)/2 \ RETURN
574 G(3)=A \ A(3,4)=(1-A)/2
580 A(1,1)=1 \ A(3,1)=0 \ A(1,2)=1 \ A(2,2)=0 \ A(1,3)=0 \ A(2,3)=1
582 A(2,1)=B(1) \ A(3,2)=B(2) \ A(3,3)=B(3)
584 N=3 \ GOSUB 310
586 FOR J=1 TO 3 \ G(J+3)=S(J) \ NEXT J \ RETURN
1000 PRINT CHR$(29);
1002 R1=T(1)^.47619 \ GOSUB 300 \ FOR I=3 TO 6 \ R1=T(I)^.47619 \ GOSUB 300 \ NEXT I
1004 R1=T(11)^.47619 \ GOSUB 300 \ FOR I=3 TO 6 \ R1=T(I+10)^.47619 \ GOSUB 300 \ NEXT I
1006 R1=T(21)^.47619 \ GOSUB 300 \ FOR I=3 TO 6 \ R1=T(I+20)^.47619 \ GOSUB 300 \ NEXT I
1008 FOR I=1 TO 6 \ R1=0 \ GOSUB 300 \ NEXT I
1010 FOR I=1 TO 6 \ R1=F(I) \ GOSUB 300 \ NEXT I
1012 PRINT CHR$(31) \ GO TO 104
2000 FOR I=0 TO 2 \ FOR J=1 TO 4 \ D(J+I*10)=L1*(1-D)*K1/(2*A1)*F(J+I*10)
2002 NEXT J \ D(5+I*10)=L1*F/A1*P(5+I*10) \ D(6+I*10)=L1*D/A1*P(6+I*10)
2004 FOR J=1 TO 6 \ I(J+I*10)=T(J+I*10)-D(J+I*10) \ NEXT J \ NEXT I
2008 PRINT \ PRINT
2010 PRINT "DIRECT (D) OR INDIRECT (I) COMPONENT OF LUMINANCE ? "; \ INPUT #0,A$
2015 IF A$="I" THEN 2030
2020 FOR I=0 TO 2 \ FOR J=1 TO 6 \ T(J+I*10)=D(J+I*10) \ NEXT J \ NEXT I
2025 GO TO 1000
2030 FOR I=0 TO 2 \ FOR J=1 TO 6 \ T(J+I*10)=I(J+I*10) \ NEXT J \ NEXT I
2035 GO TO 1000
```

Computer program for the calculation of inter-reflections using a split-flux model.

```
DX1:CNA1.BDJ
5 DIM X(6),Y(6),C(2),C1(2),P1(6),P(26),T(26),V(2),I1(2),S(3),A(3,4),R(6),F(6)
6 DIM D(26),I(26)
7 PRINT CHR$(31);
9 PRINT CHR$(29); \ FOR I=1 TO 21 \ R1=0 \ GOSUB 300 \ NEXT I
11 F(1)=25 \ F(2)=15 \ F(3)=15 \ F(4)=25 \ F(5)=25 \ F(6)=120
14 PRINT CHR$(31)
16 PRINT  \ PRINT "ROOM DIMENSIONS (IN METRES)?"
18 PRINT  \ PRINT "LENGTH ="; \ INPUT #0,L
20 PRINT  \ PRINT "WIDTH ="; \ INPUT #0,W
22 PRINT  \ PRINT "HEIGHT ="; \ INPUT #0,H
24 A1=L*W \ K1=A1/(H*(L+W))
28 FOR I=1 TO 6
30 GOSUB 260
32 PRINT  \ PRINT "REFLECTANCE= "; \ INPUT #0,R(I)
34 PRINT  \ PRINT "CHROMATICITY X= "; \ INPUT #0,X(I)
36 PRINT TAB(13);"Y= "; \ INPUT #0,Y(I)
38 C(1)=X(I) \ C(2)=Y(I) \ GOSUB 280
40 NEXT I \ PRINT  \ PRINT
42 PRINT "NOW THE LIGHTING SYSTEM"
44 GOSUB 130
46 PRINT  \ PRINT "LUMINAIRE OUTPUT IN WATTS= ";
48 INPUT #0,O \ PRINT  \ PRINT "NO. OF LUMINAIRES= ";
50 INPUT #0,Z \ GOSUB 158
52 C1(0)=1 \ C1(1)=1 \ C1(2)=1
54 PRINT  \ PRINT "FLUX FRACTION RATIO= "; \ INPUT #0,F
56 T2=0
58 L2=L1/(1+F) \ FOR I=0 TO 2 \ GOSUB 235 \ NEXT I
70 PRINT CHR$(29);
72 R1=T(1)^.47619 \ GOSUB 300 \ FOR I=3 TO 6 \ R1=T(I)^.47619 \ GOSUB 300 \ NEXT I
74 R1=T(11)^.47619 \ GOSUB 300 \ FOR I=3 TO 6 \ R1=T(I+10)^.47619 \ GOSUB 300 \ NEXT I
76 R1=T(21)^.47619 \ GOSUB 300 \ FOR I=3 TO 6 \ R1=T(1+20)^.47619 \ GOSUB 300 \ NEXT I
88 FOR I=1 TO 6 \ R1=0 \ GOSUB 300 \ NEXT I
89 FOR I=1 TO 6 \ R1=F(I) \ GOSUB 300 \ NEXT I
90 PRINT CHR$(31)
92 IF T2<0 GO TO 104
94 PRINT  \ PRINT  \ PRINT "C :COLOURS"
96 PRINT  \ PRINT  \ PRINT "LO :LIGHT FITTING OUTPUT POWER"
98 PRINT  \ PRINT  \ PRINT "LS :TYPE OF LIGHT SOURCE"
100 PRINT  \ PRINT  \ PRINT "FFR :FLUX FRACTION RATIO"
102 PRINT  \ PRINT  \ PRINT "BZ :BZ NUMBER"
104 PRINT  \ PRINT "C,LO,LS,FFR OR BZ? "; \ INPUT #0,S$ \ T2=-1
105 IF S$="SEP" THEN 1000
106 IF S$="C" THEN 120
108 IF S$="LO" THEN 124
110 IF S$="LS" THEN 126
```

```
112 IF S$="FFR" THEN 128
114 IF S$="BZ" THEN  GOSUB 130
116 GO TO 58
120 PRINT  \ PRINT "SURFACE 1,2,3,4,5 OR 6? "; \ INPUT #0,I
121 PRINT  \ PRINT  \ PRINT "R AND X,Y FOR THE" \ GOSUB 260
122 PRINT "R= "; \ INPUT #0,R(I) \ PRINT TAB(16);"X= "; \ INPUT #0,X(I) \ PRINT TAB(16);"Y= "; \ INPUT #0,Y(I)
123 C(1)=X(I) \ C(2)=Y(I) \ GOSUB 280 \ GO TO 116
124 PRINT  \ PRINT "NEW LUMINAIRE OUTPUT IN WATTS= "; \ INPUT #0,O
125 PRINT  \ PRINT "NO. OF LUMINAIRES? "; \ INPUT #0,Z \ GOSUB 158 \ GO TO 116
126 PRINT  \ PRINT "LAMP NO.? "; \ INPUT #0,N1 \ GOSUB 160
127 FOR I=1 TO 6 \ GOSUB 190 \ NEXT I \ GO TO 70
128 PRINT  \ PRINT  \ PRINT "NEW FLUX FRACTION RATIO="; \ INPUT #0,F \ GO TO 116
130 PRINT  \ PRINT  \ PRINT "BZ NO?= "; \ INPUT #0,B1
132 IF B1=1 THEN 136 \ IF B1=2 THEN 138 \ IF B1=3 THEN 140 \ IF B1=4 THEN 142 \ IF B1=5 THEN 144
134 IF B1=6 THEN 146 \ IF B1=7 THEN 148 \ IF B1=8 THEN 150 \ IF B1=9 THEN 152 \ GO TO 154
136 D=.093096+1.1171*K1-.70802*K1^2+.23617*K1^3-.039112*K1^4+2.52120E-03*K1^5 \ RETURN
138 D=7.81130E-04+1.1483*K1-.70015*K1^2+.22823*K1^3-.037249*K1^4+2.37800E-03*K1^5 \ RETURN
140 D=.026025+.95007*K1-.51692*K1^2+.15527*K1^3-.023885*K1^4+1.46000E-03*K1^5 \ RETURN
142 D=-.027528+.92475*K1-.48374*K1^2+.14254*K1^3-.021733*K1^4+1.32410E-03*K1^5 \ RETURN
144 D=8.10370E-04+.77028*K1-.36218*K1^2+.098212*K1^3-.013993*K1^4+8.06260E-04*K1^5 \ RETURN
146 D=-.026386+.63699*K1-.27732*K1^2+.073227*K1^3-.010398*K1^4+6.03620E-04*K1^5 \ RETURN
148 D=-.037394+.54261*K1-.21378*K1^2+.053176*K1^3-7.29080E-03*K1^4+4.14580E-04*K1^5 \ RETURN
150 D=-.051811+.49576*K1-.18267*K1^2+.043435*K1^3-5.79100E-03*K1^4+3.24420E-04*K1^5 \ RETURN
152 D=-.12453+.57149*K1-.25333*K1^2+.073761*K1^3-.011629*K1^4+7.37490E-04*K1^5 \ RETURN
154 D=-.1149+.39415*K1-.093688*K1^2+.01081*K1^3-1.61790E-04*K1^4-4.60100E-05*K1^5 \ RETURN
158 L1=80*O*80*Z \ RETURN
160 IF N1=1 THEN 164 \ IF N1=2 THEN 166 \ IF N1=3 THEN 168 \ IF N1=4 THEN 170 \ IF N1=5 THEN 172 \ IF N1=6 THEN 174
164 A=.055 \ B=.055 \ M2=.1978 \ M1=.4683 \ G=0 \ RETURN
166 A=.055 \ B=.06 \ M2=.197 \ M1=.465 \ G=0 \ RETURN
168 A=.05 \ B=.047 \ M2=.216 \ M1=.495 \ G=0 \ RETURN
170 A=.06 \ B=.035 \ M2=.227 \ M1=.5 \ G=.21 \ RETURN
172 A=.075 \ B=.024 \ M2=.256 \ M1=.524 \ G=.07 \ RETURN
174 A=.07 \ B=.045 \ M2=.23 \ M1=.49 \ G=.035 \ RETURN
190 U1=.24498*T(I)+.26246*T(I+10)+.12612*T(I+20)
192 V1=.20685*T(I)+.71625*T(I+10)+.07689*T(I+20)
194 W1=.13573*T(I)+.93579*T(I+10)+.49799*T(I+20)
196 U=U1/(U1+V1+W1) \ V=1.5*V1/(U1+V1+W1) \ A2=ATN((V-.4683)/(U-.1978))
198 L2=.4683+.055*SIN(A2) \ L2=(L2/V-1)/(1-L2/.4683)
200 E1=B*SIN(A2)*SIN(G)+A*COS(A2)*COS(G)+M2
201 E2=B*SIN(A2)*COS(G)-A*COS(A2)*SIN(G)+M1
202 V=E2/(1-L2*E2/M1+L2) \ U=(E1-M2)*L2*V/M1+E1
204 U1=U*1.5*V1/V \ W1=(1-U-(V/1.5))*1.5*V1/V
206 T(I)=4.9425*U1-.2201*V1-1.2178*W1
208 T(I+10)=-1.6069*U1+1.8205*V1+.1259*W1
210 T(I+20)=1.6726*U1-3.3611*V1+2.1035*W1
```

```
212 IF T(I)<0 THEN T(I)=0
214 IF T(I+10)<0 THEN T(I+10)=0
216 IF T(I+20)<0 THEN T(I+20)=0
218 RETURN
235 P1(I)=(W*(P(1+I*10)+P(2+I*10))+L*(P(3+I*10)+P(4+I*10)))/(2*(L+W))
237 I1(I)=(D*P(6+10*I)+(1-D)*P1(I)+(F*P(5+10*I)))/(2+(2/K1)*(1-P1(I))-P(5+10*I)-P(6+10*I))
239 T(5+10*I)=((L2*C1(I)/A1)*(F+I1(I))*P(5+I*10))
241 T(6+10*I)=((L2*C1(I)/A1)*(D+I1(I))*P(6+I*10))
243 V(I)=L2*C1(I)*(K1/(2*A1))*(1-D+(2*I1(I)/K1))
245 FOR K=1 TO 4 \ T(K+10*I)=(V(I)*P(K+10*I)) \ NEXT K \ RETURN
260 IF I=1 THEN 262 \ IF I=2 THEN 264 \ IF I=3 THEN 266 \ IF I=4 THEN 268 \ IF I=5 THEN 270 \ GO TO 272
262 PRINT  \ PRINT "FACING WALL (1) "; \ RETURN
264 PRINT  \ PRINT "WALL BEHIND YOU (2) "; \ RETURN
266 PRINT  \ PRINT "RIGHT HAND WALL (3) "; \ RETURN
268 PRINT  \ PRINT "LEFT HAND WALL (4) "; \ RETURN
270 PRINT  \ PRINT "CEILING (5) "; \ RETURN
272 PRINT  \ PRINT "FLOOR (6) "; \ RETURN
280 A(1,1)=1 \ A(1,2)=1 \ A(1,3)=1
282 A(2,1)=1.7765-C(1)/.349 \ A(2,2)=.5497-C(1)/.584
284 A(2,3)=2.4603-C(1)/.063
286 A(3,1)=1-C(2)/.349 \ A(3,2)=1-C(2)/.584 \ A(3,3)=1-C(2)/.063
288 A(1,4)=R(I) \ A(2,4)=0 \ A(3,4)=0
290 GOSUB 320 \ P(I)=S(1)*4.83513 \ P(I+10)=S(2)*1.39594 \ P(I+20)=S(3)*13.0178 \ RETURN
300 R4=INT(R1/32) \ R2=R1+96-(32*R4) \ R4=R4+32
302 PRINT CHR$(R4); \ PRINT CHR$(R2); \ RETURN
320 FOR K=1 TO 2 \ FOR J=K+1 TO 3 \ GOSUB 340 \ NEXT J \ NEXT K
322 S(3)=A(3,4)/A(3,3) \ FOR K=1 TO 2 \ GOSUB 345 \ NEXT K \ RETURN
340 M=A(J,K)/A(K,K) \ FOR J1=K TO 4 \ A(J,J1)=A(J,J1)-A(K,J1)*M \ NEXT J1 \ RETURN
345 M=A(3-K,4) \ FOR J=0 TO K-1 \ M=M-A(3-K,3-J)*S(3-J) \ NEXT J
347 S(3-K)=M/A(3-K,3-K) \ RETURN
400 T(1)=1698.14 \ T(2)=1153.37 \ T(3)=1729.06 \ T(4)=2142.17 \ T(5)=1662.95 \ T(6)=1344.98
402 T(11)=849.571 \ T(12)=712.424 \ T(13)=162.528 \ T(14)=1143.16 \ T(15)=1010.09 \ T(16)=310.933
404 T(21)=360.494 \ T(22)=562.89 \ T(23)=113.702 \ T(24)=672.358 \ T(25)=748.246 \ T(26)=67.1912
406 GO TO 70
410 F(1)=24.08 \ F(2)=16.0533 \ F(3)=8.02667 \ F(4)=12.04 \ F(5)=51.1875 \ F(6)=87.75 \ GO TO 70
412 F(1)=12.04 \ F(2)=8.02667 \ F(3)=16.0533 \ F(4)=24.08 \ F(5)=51.1875 \ F(6)=87.75 \ GO TO 70
414 F(1)=12.04 \ F(2)=8.02667 \ F(3)=16.0533 \ F(4)=24.08 \ F(5)=51.1875 \ F(6)=100 \ GO TO 70
416 F(1)=24 \ F(2)=8 \ F(3)=16 \ F(4)=33 \ F(5)=51.1875 \ F(6)=95 \ GO TO 70
418 F(1)=33 \ F(2)=16 \ F(3)=8 \ F(4)=24 \ F(5)=51.1875 \ F(6)=95 \ GO TO 70
1000 FOR I=0 TO 2 \ FOR J=1 TO 4
1002 D(J+I*10)=L2*C1(I)*(K1/(2*A1))*(1-D)*P(J+I*10)
1004 NEXT J
1005 D(5+I*10)=(L2*C1(I)/A1)*F*P(5+I*10)
```

```
1007 D(6+I*10)=(L2*C1(I)/A1)*D*P(6+I*10)
1009 FOR J=1 TO 6 \ I(J+I*10)=T(J+I*10)-D(J+I*10) \ NEXT J \ NEXT I
1012 PRINT  \ PRINT
1014 PRINT "DIRECT (D) OR INDIRECT (I) COMPONENT OF ILLUMINANCE ?";
1016 INPUT #0,A$
1018 IF A$="I" THEN 1030
1020 FOR I=0 TO 2 \ FOR J=1 TO 6 \ T(J+I*10)=D(J+I*10) \ NEXT J \ NEXT I
1025 GO TO 70
1030 FOR I=0 TO 2 \ FOR J=1 TO 6 \ T(J+I*10)=I(J+I*10) \ NEXT J \ NEXT I
1035 GO TO 70
```

APPENDIX 5

The prediction of perceived colour.

Introduction

The environment is subject to wide changes, both in the type and level of illumination. The human visual system has a capacity to respond to these changes, extracting useful information from images with luminance levels ranging from 10^{-6} to 10^{6} cd/m^2. Psycho-physical measurements have shown that a given change in the luminance of a stimulus often produces a relatively small change in response; the ability to discriminate relative brightness levels remaining approximately constant over a wide range of luminance. The relative invariance in the appearance of illuminated objects has led to the use of the term constancy[1]. Perceived shifts in response to changes in the colour of illumination also tend to be relatively small, a phenomenon sometimes called colour constancy. The degree to which constancy is maintained is known to depend on the visual context.

In order to describe the perception of visual stimuli in a particular context, stimuli may be categorised according to certain modes of appearance[2]. For interior lighting design the most important modes of appearance are:-

(1) The surface mode, in which the stimulus is perceived as being the result of light reflected from an object.

(2) The aperture mode, in which the stimulus provides no spatial information, being perceived as a patch of coloured light, as when a

small area of surface is viewed through a dark-walled tube.

(3) The illuminant mode, for which the stimulus appears to be a bright, self luminous body such as a fluorescent tube.

(4) The volume, or bulk mode, where the stimulus provides spatial depth cues but no surface information, appearing as a coloured transparent medium.

Constancy is generally promoted by the surface and volume modes. In the surface mode the observer is able to judge, usually quite accurately, the reflectance and colour of a surface separately from the level and colour of illumination. Lynes (1971)[3] describes the conditions which, in the context of lighting design, tend to promote constancy. Colour constancy is one aspect of visual perception, the mechanisms controlling which are still the subject of research. A comprehensive model of colour vision would account for simultaneous and successive contrast effects, the laws of colour mixture and matching, chromatic and brightness adaptation apparent hue and saturation changes for different spectral stimuli, the difference between brightness and luminance and the variation of apparent hue and saturation with changes of luminance, as well as constancy effects.

Historically, two theories of colour vision have retained consistent support, the tri-receptor theory usually attributed to Young and Helmholtz and the opponent colour model proposed by Hering. Both theories have survived because each gives a satisfactory prediction of aspects of visual perception. Physiological studies of the structure of the retina provide evidence that at a photo-chemical level there are three classes of cone receptor cell, each sensitive to a different range of wavelength within the visual spectrum. There are approximately 7 million cone cells in the retina in addition to 150 million rod cells, but only $\sim$ 1 million nerve fibres within the optic nerve, suggesting that the receptors must have their outputs encoded in some way before leaving the retina. A straight-forward three-receptor link between retina and brain is clearly not possible. The structure of the retina is known to be complex; neural interconnection between groups of cells involves signal mixing, amplification and feedback[4]. There is evidence that the relationship between some nervous responses and excitation becomes non-linear before the neural signals leave the retina. Similar and dissimilar cone cells as well as rods and cones are known to interact.

Some recent physiological studies have involved neural measurements outside the retina. Zeki (1980)[5] investigated responses of individual cells within the visual cortex of the Rhesus monkey. Cells with both spectrally opponent and non-opponent responses were found. In a striking demonstration of the specialisation of cells in the V4 area of the visual cortex, Zeki was able to identify individual cells which, when the eye was exposed to a complex image, would respond to a red surface illuminated by white light, but not to a red surface illuminated by red light. Other cells were isolated which exhibited similar specialisation for red and green stimuli.

Physiological evidence, therefore gives support to both trichromatic non-opponent and opponent response aspects of colour vision. The relative importance of the retina, the neural pathway to the brain and the visual cortex is still the subject of debate. Recent models describing the mechanism of colour vision are therefore tending to agree about the need to include aspects of the Young-Helmholtz and the Hering models. The situation is further complicated by evidence that the brain is able to modify visual perceptions according to experience. The shape of a coloured object for example, can change its perceived colour. As discovered by Bartleson[6], our memory of colour perception is often different from the original. There is a tendency for memory colours to be more saturated than the original perception. When viewing real objects, as opposed to laboratory test stimuli, it is probable that the brain acts as the final arbitor of the perception.

The Von Kries coefficient law

Colour constancy has often been linked with a feature of visual perception called adaptation, Helmholtz (1866)[7] was apparently aware of adaptation when he referred to the tendency for stimulated areas of the retina to become 'fatigued' compared to other areas. This concept was later formalised by Von Kries in his coefficient law[8]. The co-efficient law assumes that the visual system has three fundamental colour receptors each having a variable sensitivity which tends to decrease with increasing stimulation. Von Kries had earlier developed a persistence law, that if two colour stimuli look alike in one state of adaptation, they will continue to do so as the adaptation conditions change. This persistence law is embodied in the coefficient law, which is claimed to predict the tristimulus values of a colour which in a given state of adaptation, say that for illuminant A, will match a colour viewed in different illumination, for example daylight.

If R, G and B are the tristimulus values for a stimulus viewed in daylight, the coefficient law predicts the new values R', G', B' required to produce the same colour perception when viewed under a different illuminant, where

$$R' = K_R R$$

$$G' = K_G G$$

$$B' = K_B B$$

The coefficients K_R, K_G and K_B, which vary according to the illuminant used, are usually determined by matching of a non-selective surface colour for the two illumination conditions. The transformation therefore makes an implicit assumption that constancy applies for spectrally non selective surfaces. The chromaticity of the illuminant, and not its spectral properties, is important in determining the coefficients. The spectral distribution of the colour responses is assumed invariant, only the overall magnitudes being fixed by the coefficient values. The coefficient law therefore predicts that colour matches are preserved despite changes of chromatic adaptation.

Experimental measurements of adaptive shifts, such as those by Helson, Judd and Warren (1952)[9] and MacAdam (1961)[10] have shown less than perfect agreement with predictions using the coefficient law. For this reason, attempts have been made to modify the Von Kries formulae to improve the predictive accuracy.

Models for colour prediction based on the Von Kries coefficient law

The MacAdam non-linear model

MacAdam showed that a significant improvement in colour predictions resulted from a modified form of the three receptor Von Kries model in which the responses were assumed to be non-linearly related to the tristimulus values based on three fundamental colour receptors[10].

The non-linearity in the MacAdam model assumes a power law relationship between tristimulus value and response with a different exponent for each response, the exponents changing with the state of chromatic adaptation. The same ratio of responses is assumed to apply for matching colours. The model is consistent with the generally accepted view that for normal levels of illumination, metameric colour matches are undisturbed by changes in chromatic adaption; the Von Kries persistence law is maintained.

As developed by MacAdam, the non-linear Von Kries model can only be used to predict colour appearance when luminance remains constant. Work by Helson (1938)[11] and Helson, Judd and Warren (1952)[9] showed that the reflectance of a surface relative to that of adjacent colours can greatly affect the colour perception.

The Hunt linkage model

In measuring colour perceptions in both daylight and incandescent lighting, Hunt (1953)[12] observed that saturation was dependent on the luminance level of the adapting illuminant. To account for this phenomenon, Hunt proposed that neural linkages exist between the three colour channels which are dependent on the excitation level. The model consists of two stages, the first being a linear Von Kries adjustment, followed by a luminance-dependent linkage stage[13].

The Richards-Parks model.

In 1959, publications by Land described the perception of a wide range of colour from two-colour separation overlays[14]. The Land demonstrations stimulated interest in the study of colour perception. Judd (1960)[15] suggested that his empirical equations, developed in 1940[16], could predict the Land perceptions. This claim was later confirmed by Pearson et al (1969)[17]. In 1971, Richards and Parks[18] showed that a two-stage model, based on a non-linear development of the Von Kries coefficient law could also predict the Land observations.

The first stage of the model is a stright-forward application of the Von Kries law, with the assumption that constancy is incomplete. For the simulation of the Land effect, a 70% shift of the achromatic chromaticity from a daylight reference x_c, y_c towards the chromaticity of the real illuminant x_I, y_I, was found to give the best colour predictions. Assuming three fundamental response mechanisms R, Y and B, the Von Kries coefficients for complete constancy are defined by:

$$K_i = F_i(C)/F_i(I)$$

where F_i are tristimulus values and $i = R,Y,B$

for a 70% shift of the subjective neutral

$$K_i = F_i(C)/\left[0.7F_i(I) + 0.3F_i(N)\right] \quad i = R,Y,B$$

where $F_i(N)$ are the tristimulus values of the perceived neutral. This stage of the model predicts that the colours perceived from the Land projections are restricted to one dimension in colour space. The Von Kries coefficients are, however, assumed to change with local luminance changes such that:

$$k_i = K_i^{P_i} \qquad i = R,Y,B$$

where k_i are revised coefficients.

This luminance dependence is based on the observation by Helson (1938)[11] that spectrally non-selective surfaces, viewed in chromatic illumination exhibited a hue variation, dependent on the reflectance value. Surfaces with higher than average reflectance taking on the colour of the illuminant whilst those with reflectance below average appeared with a complimentary hue.

The exponents P_i are related to luminance contrast by the equation

$$P_i = \left[L_{av}/L_s\right]^{E_i} \qquad i = R,Y,B$$

where E_i is required to reduce the effective luminance ratio. An equation of this form is required for each of the fundamental responses, the modified transformation becoming

$$F_i(L_s C) = k_i F_i(L_s I) \qquad i = R,Y,B$$

The most appropriate fundamental response functions and values for E_i were derived using the comprehensive measurements of chromatic adaptation made by MacAdam. Of several fundamental response functions examined using plots of $\log\log(k_i)$ against $\log(L_s/L_{av})$, the functions $R(\lambda)$, $Y(\lambda)$ and $B(\lambda)$ were found to give the best approximations to straight lines with E_R, E_Y and E_B respectively 1/2, 1, 1.

where $R(\lambda) = 0.6\bar{x}(\lambda) + 0.5\bar{y}(\lambda) - 0.1\bar{z}(\lambda)$

$$Y(\lambda) = \bar{y}(\lambda)$$

$$B(\lambda) = 0.67\bar{y}(\lambda) + 0.33\bar{z}(\lambda)$$

The model was applied to the prediction of colours perceived from two-colour separation displays such as those investigated by Land (1959)[14]. The best fit to the results of Pearson et al (1969)[17] was achieved with R, Y and B fundamental responses with $E_r = 1/8$, $E_y = 1/4$ and $E_B = 1/4$. The percentage adaptation to the illuminant was also varied and 70% found to be optimum. Predictions for a quilt display similar to that used by Pearson et al proved to be at least as accurate as those made using the Helson-Judd equations. Unlike the Judd model[16], the Richards-Parks equations are not based on a particular colour spacing system and mathematically are more elegant.

The Richards one-stage model.

A further development of the Richards-Parks model[18] was suggested by Richards (1972)[19]. The new model is based on discounting the illuminant and Helson's rule of contrast, but differs in its description of the former. The modification simplifies the model mathematically and may have implications for the mechanism of colour vision.

In the original model, discounting the illuminant is assumed to be fixed at a 70% level and is described by the equation

$$f(K_i) = K_i/(0.7 + 0.3K_i)$$ where K_i is the Von Kries coefficient $i = r,y,b$

This equation is replaced by $g(K_i) = K_i^{0.7}$

Since the second stage contrast equation is also in exponent form, the two equations may be combined to give the Von Kries coefficient

$$k_i = g(K_i)^{p_i} = K_i^{0.7p_i} \qquad i = r,y,b$$

$$\text{where } P_i = (L_{av}/L_s)^{E_i}$$

The original model required different values of E_i for the three colour channels R, Y and B

for example

$$P_B = (L_{av}/L_s)^{\frac{1}{2}}$$

$$0.7P_B = (0.5\ L_{av}/L_s)^{\frac{1}{2}}$$

and

$$P_R = (L_{av}/L_s)^{1/4}$$

$$0.7P_R = (0.24\ L_{av}/L_s)^{1/4}$$

Discounting the illuminant and Helson contrast are now seen to have the same effect on the Von Kries coefficients, suggesting that the visual mechanism may be similar for both effects. A change in the discounting of the illuminant from 100% to 70% is equivalent, for the blue channel, to a 50% fall in the average adaptation level whereas for the red channel this would require a 76% fall. Thus, the effective mean level of adaptation for the blue channel is higher, a finding consistent with the results to be described in chapter 5 for adaptation to the television display.

Application of the Hunt linkage model

In 1977, Takahama et al[20] produced a two-stage model based on the empirical linkage system proposed by Hunt[13]. The first stage is a linear Von Kries sensitivity adjustment dependent on the chromaticity of the test and reference illuminants. Each of the three receptors is assumed to have a neural linkage with the other two. The contribution from the other receptors in the final output of a one colour channel is dependent on the stimulus level of the receptors associated with that channel. If a receptor is highly stimulated, the contribution from the other two channels is assumed to be small; as the stimulus is decreased, the contribution from the other channels increases. This neural inter-mixing of the colour channels forms the second stage of the model, which is able to account for the variation in colour saturation with varying luminance, as observed by Hunt (1953)[12].

The model was also tested using the results of Burnham et al[21] for chromatic adaptation at a fixed luminance level. A comparison with the predictions using the Von Kries coefficient law showed the linkage model to give a significant improvement in predictive accuracy.

A development of the MacAdam non-linear model.

The non-linear model proposed by MacAdam described above is, like the linear Von Kries model, strictly only applicable to the case in which the luminance remains constant. Having developed the linkage model to predict the effects of luminance changes on perceived saturation, Nayatani et al (1981)[22] extended the MacAdam model to include luminance variation. This new model proved to be as successful as the linkage model described above, in perceived colour prediction, adding support to the suggestion by Stiles (1976)[23] that the two models may be related.

As in the MacAdam model, the non linearity takes the form of a power law applied to the three colour responses derived from the standard Von Kries first stage. The exponents increase in value as the excitation of each channel increases and are therefore both chromaticity and luminance dependent. The model is similar to that of Richards described above, except that the non-linearity is applied to the responses rather than to the Von Kries coefficients. The Richards model was, however, developed in an attempt to predict the effects of the Land two colour projections rather than as a general description of adaptation. Whereas Richards assumes an average level of adaptation, about which relative luminance changes influence the exponents applied to the Von Kries coefficients, the Nayatani model includes luminance as an absolute value.

As well as having an historical importance, the Von Kries coefficient law has, therefore, survived to be incorporated into contemporary models describing colour perception. Models for colour prediction not based on the Von Kries law have also been developed, two examples of which are the Helson-Judd model[16] and the model by Burnham et al (1957)[21].

The Helson-Judd model

In a classical paper written in 1940[16], a set of semi-empirical colorimetric equations formulated by Helson and Judd was published. These equations are designed to predict hue, saturation and lightness of a set of Munsell chips viewed under chromatic illumination. The equations are based on the co-ordinates of r,g,b 'uniform' colour triangle.

Taking discounting of the illuminant as a starting point, the following equations are consistent with the simple Helmholtz model

$$r_n = r_f - k(r_f - r_w)$$
$$g_n = g_f - k(g_f - g_w) \qquad 1$$

where r_w, g_w are tri-linear co-ordinates of reference neutral.

Equation 1 is equivalent to the Von Kries coefficient law. The constant k represents the level of adaptation with k = 0 corresponding to complete adaptation, k = 1 to zero adaptive shift.

In order to achieve accurate predictions of his experimental measurements, Judd found it necessary to add four refinements to equation 1. The resulting equations have a more complex form:

$$r_n = r_f - D_f \left[0.1L'(r_f - 0.360) - 0.018b_f A_f (L')^2 \log_{10} 2000I\right]$$
$$g_n = g_f - D_f \left[0.1L'(g_f - 0.300) - 0.030\right] \qquad 2$$

where L' is the predicted lightness

I is the illuminance

D_f is the distance from the 'daylight point' of the adapting stimulus.

The refinements made by Judd may be summarised:

The L' term is used to account for the observation by Helson (1938)[11] that areas of lower than average reflectance tend to take on the complementary colour to that of the illuminant whilst areas with higher than average reflectance take on the colour of the illuminant, a phenomenon generally known as Helson's rule of contrast.

D_f ensures that the adaptation becomes less complete as the saturation of the stimulus is increased. For complete adaptation $D_f = 0$, this being the condition for guaranteed colour constancy.

Two other modifications of equation 1 are also included to take account of after-image effects. For high adapting field illuminances I and reflectance A_f, the after-image effects of a blue illuminant are intensified, leading to the inclusion of the product $b_f A_f \log_{10} 2000I$.

Finally, the complementary colour produced by successive contrast differs from that for simultaneous contrast in having more red-blue. This effect is accounted for by the $(L')^2$ term in the expression for r_n, the addition of $0.30D_f$ to g_r and the use of the co-ordinates 0.36, 0.30 for the reference 'white' point.

Having predicted the achromatic point r_n, g_n, the hue, saturation and lightness of other stimuli are predicted using equations:

$$\text{hue} = f\left[(r - r_n)/(g - g_n).\text{sgn}(r - r_n)\right]$$

$$\text{saturation} = 50\left[(r - r_n)^2 + (g - g_n)^2 + (b - b_n)^2\right]^{1/2}$$

$$\text{lightness} = \frac{10(A - 0.03)(A_f + 1.00)}{(1.00 - 0.03)(A_f + A)}$$

where r,g,b is the perceived chromaticity of the point of interest.

Compared with the models based on the Von Kries coefficient law, the Helson-Judd equations have the advantage of not depending on controversial data defining receptors. Unfortunately the model lacks mathematical elegance.

Pearson et al (1969)[17], successfully applied the Helson-Judd equations to the prediction of colours perceived from two and three colour, computer-generated displays.

The Burnham, Evans and Newhall model.

In a study of colour perception under changing illumination Burnham et al (1957)[21] suggested that their analytical approach could be extended to predict perceived colour under controlled viewing conditions. Using an haploscopic matching technique, one eye was adapted to a reference illuminant whilst the other was adapted to the illuminant of interest. A set of twelve different colour samples was viewed under the reference adaptation conditions whilst a variable stimulus was adjusted for match under the adaptation being studied. The results for several observers were averaged to produce twelve pairs of chromaticities for each illuminant being compared to the reference.

Unlike Judd (1940)[16], no attempt was made to describe the results in terms of colour perception phenomena. The above chromaticity pairs were used to derive a 4 x 3 transform matrix M such that

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} M \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix}$$

Using this equation it is claimed that the perceived chromaticity of any colour sample under the illuminant used to drive $[M]$ may be predicted, knowing the chromaticity of that same sample under the reference illuminant. The accuracy of the technique is therefore totally dependent on the experimental conditions used to derive matrices. If, for example, luminance level is changed a new matrix would be required.

References

1. Katz, D.
The world of colour. Kegan, Paul, Trench, Trubner. London (1935).

2. Judd, D. B.
ASTM, Special Publication, No. 297, Philadelphia (1961).

3. Lynes, J. A.
Lighting Res. and Technology. vol. 3, no. 1, p. 24, (1971).

4. Bartleson C. J.
A review of the structure of the retina is given by C. J. Bartleson. J.S.M.P.T.E., 77, 1, p. 1, (1968).

5. Zeki, S.
Nature, 284, p. 412, (April 1980).

6. Bartleson, C. J.
J. Opt. Soc. Am., 50, p. 73, (1960)

7. Helmholtz, H. v.
Die Lehre von den Gesichtesempfindungen Handbuch der Physiol., Opt. II, Hamburg & Leipzig, Leopold Voss, 1911, (1st ed. 1866).

8. Von Kries, J.
Arch. Anat. Physiol., (Leipzig. Physiol. Abt.), 2, p. 503, (1878).

9. Helson, H., Judd, D. B. and Warren, M.
Ill. Eng., 47, p. 221, (1952).

10. MacAdam, D. L.
Vision Res., 1, p. 9, (1961).

11. Helson, H.
J. Exptl. Psychol., 23, p. 439, (1938).

12. Hunt, R. W. G.
J. Opt. Soc. Am., 43, p. 479, (1953).

13. Hunt, R. W. G.
Visual problems of colour. p. 639, (1957), NPL Symposium No. 8 (Her Majesty's Stationery Office, London, 1958).

14. Land, E. H.
Proc. Natl. Acad. Sci (U.S.), 45, 115, p. 636 (1959).

15. Judd, D. B.
J. Opt. Soc. Am., 50, p. 254 (1960).

16. Judd, D. B.
J. Opt. Soc. Am., 30, p. 2, (1940).

17. Pearson, D. E., Rubinstein, C. B. and Spivack, G.J.
J. Opt. Soc. Am., 59, p. 644, (1969).

18. Richards, W. and Parks, E. A.
J. Opt. Soc. Am., 61, 7, p. 971, (1971).

19. Richards, W.
J. Opt. Soc. Am., 62, 5, p. 697, (1972).

20. Takahama, K. Sobagaki, H. and Nayatani, Y.
J. Opt. Soc. Am., 67, 5, p. 651, (1977).

21. Burnham, R. W., Evans, R. M. and Newhall, S.M.
J. Opt. Soc. Am., 47, p.35. (1957).

22. Nayatani Y, Takahama, K. and Sobagaki, H.
Color Res. & Appl., 6, no. 3, (Fall 1981).

23. Stiles, W. S.
J. Colour Group (London), no. 11, p. 105, (November 1976).

Appendix 6

APPLICATIONS OF THE SYSTEM

The display system was originally conceived as a hardware/ software package which would assist students of architecture to understand quickly and easily some fundamental concepts of lighting and colour.

In an educational context, the system has proved valuable for the reinforcement of the lighting concepts, flux fraction ratio, BZ classification and direct ratio which allow the selection of lighting installations from manufacturers' data. The display is interactive, permitting the result of design changes to be seen without delay thus giving, for example, a clear illustration of the trade-off between illumination and reflectance, between direct and indirect lighting. In addition to reflectance differences, the student may examine alternative colour combinations and how colour harmony may relate to room appearance. By providing a wide range of colours the display offers a means of illustrating the colour parameters hue, value and chroma of the Munsell system, BS hue, greyness and weight, and CIE chromaticity.

The display also provides a very clear illustration of the problems associated with the colour rendering of different light sources; that colour appearance is not simply related to the colour of the light source.

The prototype display used in the Plymouth Polytechnic School of Architecture has produced a favourable response from both staff and students. The system has also proved an aid to the teaching of colour science in the Plymouth Polytechnic course in scientific and technical graphics, providing a convenient, calibrated source of colour for demonstrating the significance of reflectance, chromaticity and the Munsell system.

In addition to its educational role, the display system may have potential in a broader context, enabling lighting design to be carried out without the need to understand concepts such as flux, luminance and chromaticity. A lighting system and colour scheme could be chosen with reference to manufacturers' catalogues, the end product being seen without the use of text-books or additional information. If the light distribution does not meet the I.E.S. standards for task illuminance, wall/task ration or ceiling/task ratio, immediate changes can be made to achieve an acceptable design. Finally, the system would allow the designer to experiment with the effect of different light sources on surface colour appearance.

At a commercial level, the system has stimulated interest from paint manufacturers, who recognise the potential of being able to simulate the effects of inter-reflection, colour harmony and the colour rendering of a wide range of light sources on surface appearance.

Originally developed as an educational aid, the system was designed to allow instantaneous image changes, for this reason the image generation has a strong hardware bias.

As with most hardware based systems, specific performance objectives are achieved at the expense of versatility. In a more general design context, the need may be for a more realistic and detailed image, with the ability to make rapid changes of lesser importance. The two applications therefore, have conflicting requirements for the display design. In the case of paint retailing, a compromise, involving a small increase in image detail, for example, gradation of luminance, may be preferred.