UDC 681.3.048

Ref. No. : ISO/R 1831-1971 (E

ISO

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION

ISO RECOMMENDATION

R 1831

PRINTING SPECIFICATIONS

FOR OPTICAL CHARACTER RECOGNITION

1st EDITION

November 1971

Printed in Switzerland

Also issued in French and Russian. Copies to be obtained through the national standards organizations.

BRIEF HISTORY

The ISO Recommendation R 1831, *Printing specifications for optical character recognition*, was drawn up by Technical Committee ISO/TC 97, *Computers and information processing*, the Secretariat of which is held by the American National Standards Institute (ANSI).

Work on this question led to the adoption of Draft ISO Recommendation No. 1831, which was circulated to all the ISO Member Bodies for enquiry in March 1969.

The Draft has been approved, subject to a few modifications of an editorial nature, by the following Member Bodies :

Czechoslovakia	Italy	Switzerland
Denmark	Japan	Thailand
France	Korea, Rep. of	Turkey
Germany	Netherlands	United Kingdom
Greece	Spain	U.S.S.R.
Israel	Sweden	

The following Member Bodies opposed the approval of the Draft :

Belgium
New Zealand
U.S.A.

This Draft ISO Recommendation was then submitted by correspondence to the ISO Council which decided to accept it as an ISO RECOMMENDATION.

CONTENTS

ISO Recommendation R 1831 November 1971

PRINTING SPECIFICATIONS FOR OPTICAL CHARACTER RECOGNITION

1. GENERAL

1.1 Purpose

The purpose of this ISO Recommendation is to provide first guidelines upon which industrial standards could be based for paper and printing to be used in optical character recognition (OCR) systems.

It defines tentatively the relevant parameters and their measurement methods. Field experience has shown that more precise definitions and methods must be developed. Appropriate work is under way. Clauses under review are identified in Appendix Z.

1.2 Scope

This ISO Recommendation contains basic definitions, measurement requirements, specifications and recommendations for OCR paper and print. Additional restrictions will often need to be imposed and additional pertinent variables will need identification and control. Such items as document size, the mechanical properties of the paper, the degree of control necessary over possible variations, and the format details of the particular application should be resolved by those concerned.

This ISO Recommendation applies to the character sets given in ISO Recommendation R 1073, *Alphanumeric character sets for optical recognition.*

Three major parameters of a printed document for OCR media are covered, namely :

(*a*) optical properties of the paper to be used;

(*b*) optical and dimensional properties of the ink patterns forming OCR characters;

(*c*) basic requirements related to the position of OCR characters on the paper.

The major factors of each of these areas pertinent to OCR are identified. Definitions of these items are given and bases for measurements are established.

Basic specifications applicable to all OCR materials are imposed and recommendations for the implementation of an OCR system are made.

Because of the widely divergent nature of OCR applications this ISO Recommendation does not include all of the necessary or prudent specifications or considerations that may be necessary for a successful OCR system.

1.3 Interpretation of the Recommendation

The values in this ISO Recommendation represent the specifications for supplies and the limit of performance for a printing system to be used for the preparation of OCR media. They are established on the basis that they are reasonably obtainable. However, it must be recognized that many parameters are subject to variation, and deviations from the specified limits may occur.

The degree to which these deviations are allowed (in cases where the specification is not already expressed in statistical terms) will depend upon the specific application and should be evaluated by the users and suppliers before a system is to be established.

Furthermore, although the limit of each parameter is given as an independent variable, a deterioration in reader performance is likely if the limits of more than one parameter are approached simultaneously. Every effort should be made to keep well within the limits.

It is not unknown for there to be a deterioration in print quality during the time which elapses between printing and OCR processing. Such changes are difficult to measure and this ISO Recommendation makes no distinction between the state of OCR material immediately after printing and the state immediately before reading.

1.4 Use of the Recommendation

In using and referring to this ISO Recommendation in any particular application it is necessary to specify the selection from a number of choices so that the proper portions of the document can be applied. These choices are selection of the font, the font size, the character repertoire, the spectral characteristics of the paper and printed images, the paper opacity, and the strokewidth tolerances.

2. SPECTRAL REQUIREMENTS

2.1 General

This section contains the definition of spectral bands of interest for OCR applications.

They must be defined since character readers operate in specific spectral regions and paper and ink characteristics change with the wavelength considered.

2.2 Spectral bands (see Appendix Z)

In this clause a set of bands is defined as reference for the paper and printed image specification. Their use and the measuring procedures are specified in the clauses on paper reflectance (3.2), paper opacity (3.4) and PCS measurement (4.7.2).

Band	Peak in nm	Bandwidth in nm 50 % level
B 400	See below	
B 425	425 ± 5	50 or less
B 460	460 ± 5	60 or less
B 490	490 ± 5	60 or less
B 530	530 ± 5	60 or less
B 570	570 ± 10	100 or less
B 620	620 ± 10	100 or less
B 680	680 ± 10	120 or less
B 900	900 ± 50	400 or less

The bands B 425 up to B 900 represent the spectral responses required from the complete measuring instrument (light source, filter, detector). These responses must be smooth curves without secondary peaks and with no major parts of the response curve beyond the specified 50 % points. The energy content of the illumination at wavelengths shorter than 400 nm should not exceed 5 % of that in the particular band under consideration.

The shortest wavelength band, B 400, is defined somewhat differently (see Appendix Y, clause Y.1.1.3), as follows :

> The light source must have a peak output at 400 ± 10 nm with a bandwidth of not more than 60 nm. The detector must have a uniform response (not less than 75 % of the peak response) over the range 365 to 500 nm.

3. PAPER SPECIFICATIONS FOR OCR

3.1 General (see Appendix Z)

The papers to be used in OCR applications should be white (see Appendix Y, clause Y.1.1.2), have a flat finish and low gloss and should be of high opacity.

Fluorescent additives should be avoided. Paper for OCR should also be free from watermarks and coloured patterns.

3.2 Paper reflectance

The measurements in this section deal only with diffuse reflectance. The reflected light used for measurement shall exclude specularly reflected light.

Unless otherwise specified, all reflectance values are referred to magnesium oxide (MgO) as the primary white standard. The reflectance of MgO is the 100 % value. Absence of any light of the wavelengths of interest is the 0 % value.

The average paper reflectance measurements shall be made using the infinite pad method, i.e. the samples being measured must have a backing of a sufficient number of paper thicknesses of the same type such that doubling the number will not change the measured value of reflectance.

The variation in paper reflectance shall be measured using the black backing method, i.e. the sample being measured must be backed with black of not more than 0.5 % reflectance.

3.2.1 *Average reflectance*

3.2.1.1 MEASUREMENT AREA. Each measurement of average reflectance shall be made using an area of at least 65 mm^2 (0.1 in^2). The area will be in the shape of a circle or of a regular polygon.

3.2.1.2 VISUAL SPECTRUM. The average reflectance of the paper shall not be less than 60 % in the range from 425 to 500 nm and shall not be less than 70 % in the range from 500 to 700 nm.

Average reflectance may be determined either by means of spectrophotometric measurements or, by a number of reflectance measurements in different spectral bands.

For white papers and slightly but uniformly coloured papers it is sufficient to measure the reflectance in the two following spectral bands :

- B 425;
- B 530 or B 570 or any band peaking in between and having a bandwidth smaller than or equal to 100 nm. (The CIE/Y spectral energy distribution also referred to as "photopic luminosity function" satisfies this requirement.)

In doubtful cases where these two band measurements may not establish the required reflectance throughout the whole range, it is necessary to make reflectance measurements in a greater number of bands.

The following set of bands may be used for the purpose :

B 425, B 460, B 490, B 530, B 570, B 620, B 680.

Any other choice of bands may be employed provided they adequately cover the visible spectrum.

When the near infra-red (IR) spectrum is of interest, an average reflectance of 70 % in the band B 900 is required. Since white and slightly coloured papers which meet the previous specifications will usually present an average reflectance greater than 70 % in the near infra-red spectrum, reflectance measurements in this band usually are not necessary.

In cases where the near ultra-violet (UV) spectrum is considered, the average reflectance should be greater than 55 % when measured in the B 400 band. White papers will usually meet this requirement.

3.2.2 *Variation in paper reflectance.* Variation in paper reflectance is defined as the standard deviation of reflectance measurements, taken over well separated circular areas of diameter 0.2 mm (0.008 in) : see Appendix Y, clause Y.1.2.

Two classes of variations in paper reflectance are specified, namely :

– standard deviation $\leqslant$ 3.5 % of the mean reflectance (for high opacity paper : see clause 3.4.3.1);

– standard deviation $\leqslant$ 5 % of the mean reflectance (for medium opacity paper : see clause 3.4.3.2).

The specification on variation in paper reflectance must be satisfied in the following bands :

– B 425;

– B 530 or B 570 or any band peaking in between and having a bandwidth smaller than or equal to 100 nm. (The CIE/Y spectral energy distribution satisfies this requirement);

– B 900.

In practice the measurements may usually be limited to the most critical band.

In doubtful cases where a single band measurement may not be sufficient to show that the specification is satisfied throughout the whole spectrum it is necessary to use the three bands.

3.3 Dirt in paper

The dirt count in paper may not exceed 10 parts per million as determined by TAPPI (Technical Association for the Pulp and Paper Industry, 360 Lexington Avenue, New York, N.Y., USA) method T 437 - ts - 63.

All foreign material 0.01 mm^2 (0.000 012 in^2) and larger shall be counted.

3.4 Paper opacity (see also Appendix Y, clause Y.1.3)

3.4.1 *Definition of paper opacity.* Opacity (paper backing) is the ratio (expressed as a percentage) of the average reflectance of a specimen backed with black of not more than 0.5 % reflectance, to the average reflectance of the same specimen backed with an infinite pad.

3.4.2 *Measurement of paper opacity.* Paper opacity shall be measured using B 530 or B 570 or any band peaking in between and having a bandwidth smaller than or equal to 100 nm. (The CIE/Y spectral energy distribution satisfies this requirement.)

In choosing the class of paper opacity it is important that the recommendations given in Appendix Y, clause Y.1.3.2 be considered.

3.4.3 *Classes of opacity.* Papers acceptable for OCR fall into two classes, based on opacity.

3.4.3.1 HIGH OPACITY PAPER, has an opacity of not less than 85 %.

3.4.3.2 MEDIUM OPACITY PAPER, has an opacity of at least 65 % but less than 85 %.

4. CHARACTERISTICS OF THE PRINTED IMAGE (see Appendix Z)

4.1 General

This section contains specifications and quality control criteria pertaining to individual OCR characters and marks, i.e. without consideration of the relationship between the individual printed image of an OCR character and any other printing on a document. Relevant specifications for this relationship are contained in section 5.

The specifications in sections 4 and 5 pertain to printed images and not to type faces.

The performance of OCR systems depends to a large extent on the print quality. Hence, every effort should be made to provide "good" print quality. i.e. :

The printed character should present as high a contrast as possible to the background document.

Strokewidths should be held as close as possible to the nominal.

There should be no voids within the stroke outline. When this cannot be prevented the number and size of individual voids should be minimized and the distance between them should be as great as possible.

There should be no extraneous ink within the clear area. When this cannot be prevented the number and size of individual spots should be minimized and the distance between them should be as great as possible.

The mean shape centreline of the printed image should be held as close as possible to the nominal. Since variations can seriously affect reading performance, type designers and print device manufacturers are cautioned to take care and use techniques which produce printed images conforming to this ISO Recommendation.

In order to achieve the print quality required for OCR it should be understood that in comparison with non-OCR applications special precautions, including adjustments and maintenance, may have to be taken and the ribbon life of impact printers will usually be shortened.

The reflectance specifications in this section deal only with diffuse reflectance. The reflected light used for measurement shall exclude specularly reflected light.

Unless otherwise specified, all reflectance values are referred to magnesium oxide (MgO), as the primary white standard. The reflectance of MgO is the 100 % value. Absence of any light of the wavelength of interest is the 0 % value.

The reflectance measurements shall be made using the infinite pad method, unless otherwise specified. The sample being measured must have a backing of a sufficient number of paper thicknesses of the same type such that doubling that number will not change the measured value of reflectance. There should be a good understanding of the spectral properties of the ink, paper and the OCR scanners used. Where the spectral properties of the reader are not known, it is recommended that inks with a high absorption in all bands, from B 400 to B 900 inclusive, be used, e.g. carbon black pigment inks. However, it should be recognized that certain printers cannot use carbon black inks, and therefore their printing will have a high absorption only in the visual spectrum.

The requirement of human legibility imposes the use of inks which have good absorption in the visible range even where near infra-red and near ultra-violet spectrums are used for machine reading.

4.2 Measuring gauge

Printed images are measured using a gauge showing the minimum and maximum character outline limits (COL).

These limits are constructed by superimposing the minimum and maximum strokewidths, as specified on the following page, symmetrically about each point of the character centreline drawing (see Appendix Y, clause Y.2.2).

Before making measurements, the printed image must be visually aligned to give "best fit" with the gauge.

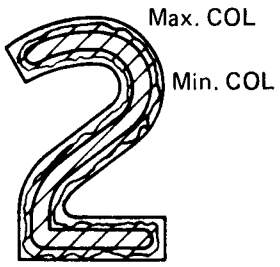


FIGURE 1 – Gauge in its "best fit" position

4.3 Mean stroke edge

Mean stroke edge is defined as the integrated average of the edge irregularities estimated visually along any length of 0.6 mm (0.024 in), parallel to the COL. Mean stroke edges of a character must be contained between maximum and minimum COL.

Dimensions in millimetres

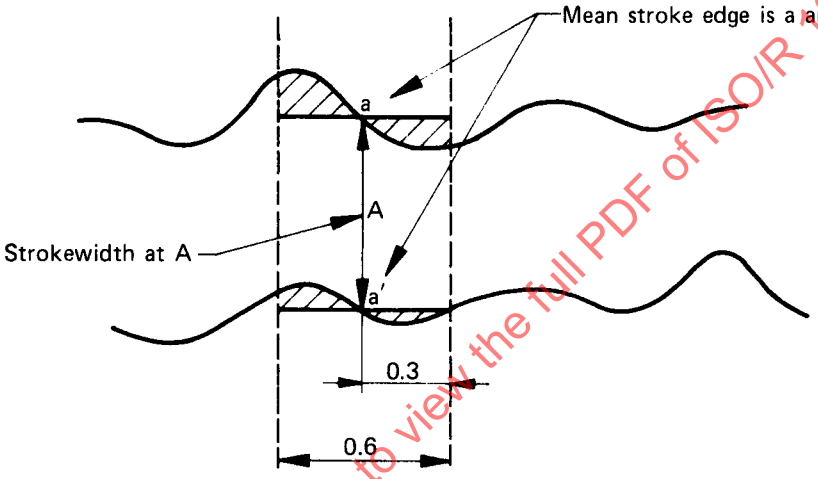


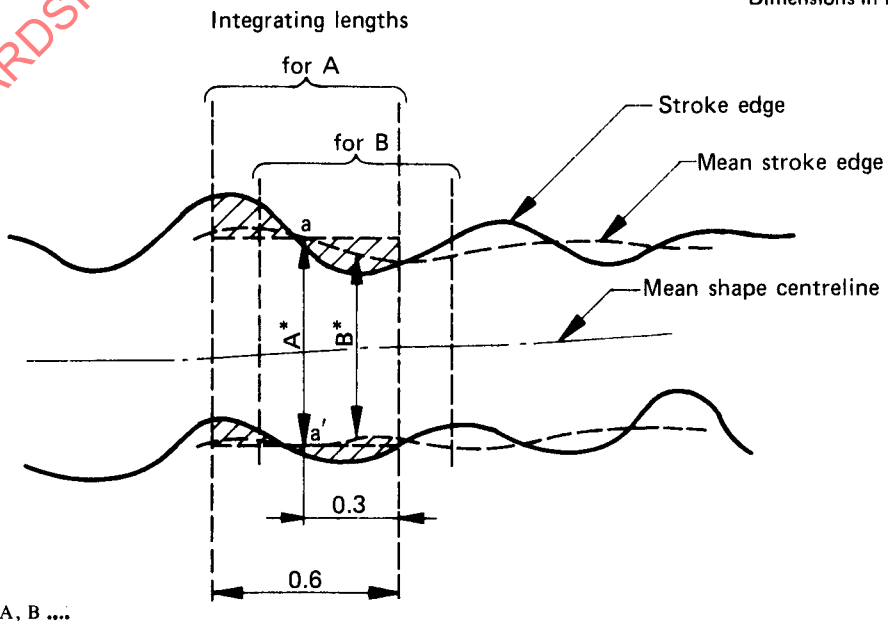
FIGURE 2

4.4 Centreline deviations

The assembly of the smoothed centrelines of the actual printed strokes is called the mean shape centreline.

The distance between any two points on the mean shape centreline of the printed image shall not differ by more than 0.075 mm (0.003 in) from the nominal distance between the equivalent two points on the ideal centreline.

Dimensions in millimetres



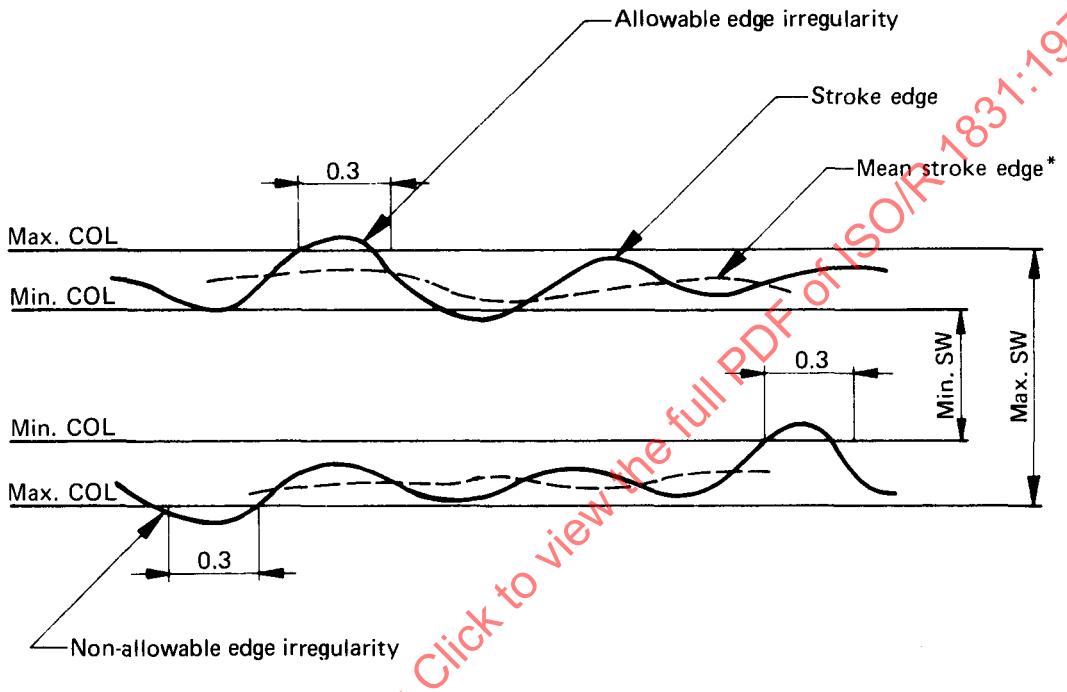
* Strokewidth is A, B ....

FIGURE 3

4.5 Strokewidth

Strokewidth is the distance between mean stroke edges measured perpendicular to the mean shape centreline. The variations in strokewidth to be expected in the normal printed output will differ according to the type of printing device employed.

Dimensions in millimetres



* The mean stroke edge is between Max. and Min. COL

FIGURE 4

Two ranges of strokewidth can be identified, a small range X and a larger range Y.

Range X can be tolerated without significant deterioration in reading performance.

As strokewidth extends beyond range Y, the reader performance may degrade rapidly.

It is expected that most printing devices for OCR will produce average printing with strokewidths in a range larger than X but smaller than Y. On the other hand, there are specially controlled printing devices and printing processes which can conveniently and economically produce average printing with strokewidth within range X.

The nominal strokewidths and the tolerances about them are as follows :

O C R – A

Size	Nominal strokewidth		Range X		Range Y	
	mm	in	mm	in	mm	in
I	0.35	0.014	± 0.08	± 0.003	± 0.15	± 0.006
II	0.35	0.014	± 0.08	± 0.003	± 0.15	± 0.006
III	0.38	0.015	± 0.08	± 0.003	± 0.18	± 0.007
IV	0.51	0.020	± 0.13	± 0.005	± 0.25	± 0.010

O C R – B

Size	Nominal strokewidth		Range X		Range Y	
	mm	in	mm	in	mm	in
I	0.35	0.014	± 0.08	± 0.003	± 0.15	± 0.006
	0.31*	0.012*	± 0.08	± 0.003	+ 0.19 – 0.11	+ 0.008 – 0.004
II	0.35	0.014	± 0.08	± 0.003	± 0.15	± 0.006
	0.31*	0.012*	± 0.08	± 0.003	+ 0.19 – 0.11	+ 0.008 – 0.004
III	0.38	0.015	± 0.08	± 0.003	± 0.18	± 0.007
	0.34	0.013*	± 0.08	± 0.003	+ 0.22 – 0.14	+ 0.008 5 – 0.005 5

* The strokewidth tolerances given in the table apply only to the following characters among the set of characters having nominal strokewidth 0.31 mm (0.012 in) for sizes I and II, or 0.34 mm (0.013 in) for size III :

£ $: ; < % > ? [! # & ¤] (=) –

For the remaining characters of nominal strokewidth 0.31 mm (0.012 in) for sizes I and II, or 0.34 mm (0.013 in) for size III, see Appendix Y, clause Y.2.9.

For numeric applications the tolerance range for size III, range Y, may have to be widened. The following limits shall not be exceeded :

+ 0.28 mm (0.011 in)
– 0.18 mm (0.007 in)

and the distance between the mean edges of parallel adjacent strokes shall not be less than 0.2 mm (0.008 in).

4.6 Edge irregularities

Any extension of the stroke edge outside the maximum COL should not exceed 0.3 mm (0.012 in), measured visually along the maximum COL (see Figure 4).

Any extension of the stroke edge inside the minimum COL should not exceed 0.3 mm (0.012 in), measured visually along the minimum COL. Edge irregularities must also meet the specifications on spots and voids (see clauses 4.8 and 4.9).

4.7 Print contrast signal

The contrast between a printed image and the paper on which it is printed is described by means of the print contrast signal (PCS).

4.7.1 *Definition of PCS.* The PCS is defined by the equation

$$PCS_p = \frac{R_w - R_p}{R_w}$$

where

R_w is the maximum reflectance found within the area of interest to which the PCS of point p is referenced. (In measuring printed images, this area of interest should be a rectangle approximately twice the nominal character height by twice the nominal character width and centred on the character being measured);

R_p is the reflectance at p.

The reflectances R_w and R_p are measured within a circular area 0.2 mm (0.008 in) diameter.

The PCS requirements that apply to the printed image are stated in clauses 4.11, 4.12 and 4.13.

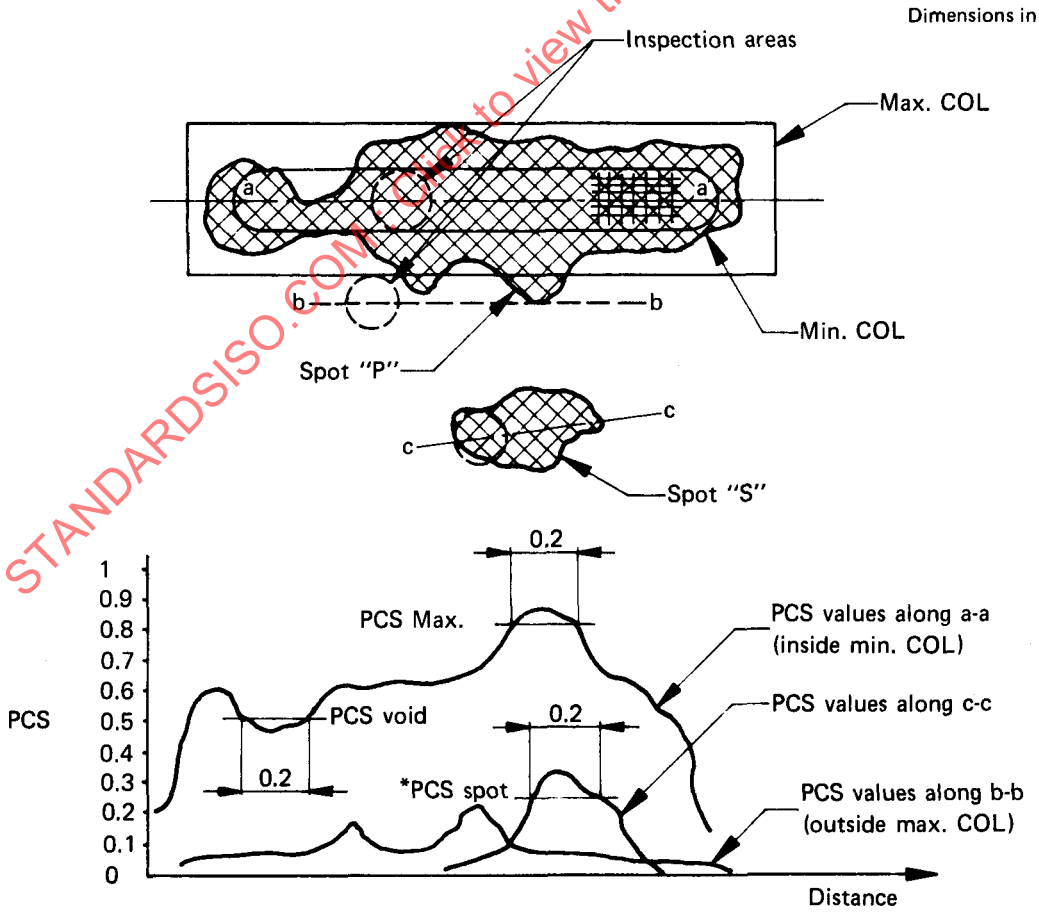
4.7.2 *Measurement of PCS.* The specification for PCS must be met in one or more of the following bands :

– near ultra-violet : B 425;

– visual : B 530 or B 570 or any other band peaking in between and having a bandwidth not exceeding 100 nm. (The CIE/Y spectral energy distribution satisfies this requirement);

– near infra-red : B 900.

The particular band(s) chosen will depend upon the characteristics of the reading and printing equipment in the system. It is important that the recommendations given in Appendix Y, clause Y.2.6, be considered.



* If spot "S" were not present the PCS spot level would be determined by spot "P". The character represented is a minus sign.

FIGURE 5

4.8 Spots

4.8.1 *Definition.* Spots are defined as areas outside any maximum COL, which can be identified as contrasting with the background. They can be measured either visually or in terms of PCS. Both methods constitute an acceptable procedure for measurement of the printed image for most printing. However, where there is a substantial question of subjective visual judgment the instrumental method based on PCS must be used.

Character associated spots are those spots within a rectangle, twice the character height and twice the character width, centred on the character. For adjacent characters, spots should be related to the nearest character.

Spots remote from characters are discussed in Appendix Y, clause Y.2.8.

4.8.2 *PCS measurements of spots.* Sizes of spots depend on the PCS level at which they are measured. For a specific PCS level (see Appendix Y, clause Y.2.7.1), a spot is described as "allowable" if it satisfies the following conditions :

(1) The distance the PCS measuring aperture can be moved in a straight line, so as to give a PCS above the chosen level, nowhere exceeds 0.2 mm (0.008 in). The measuring aperture must remain at all times outside the maximum COL.

(2) The distance, centre to centre, of the spot from the nearest other spot, detected at the same PCS level, is at least 1.0 mm (0.040 in).

"PCS spots" is defined as the minimum PCS level at which all character associated spots are allowable (see Figure 5).

4.8.3 *Visual measurement of spots.* A spot is allowable if it satisfies the following conditions :

(1) It can be contained entirely within a circle of 0.2 mm (0.008 in) diameter, estimated visually.

(2) Its distance, centre to centre, from the nearest other spot is at least 1.0 mm (0.040 in).

Small spots or groups of small spots which are contained in a circle of 0.2 mm (0.008 in) diameter are allowed in unlimited number if the total area of the spot(s) contained in the circle of 0.2 mm (0.008 in) diameter is smaller than one-third of the area of the circle, (see Figure 6).

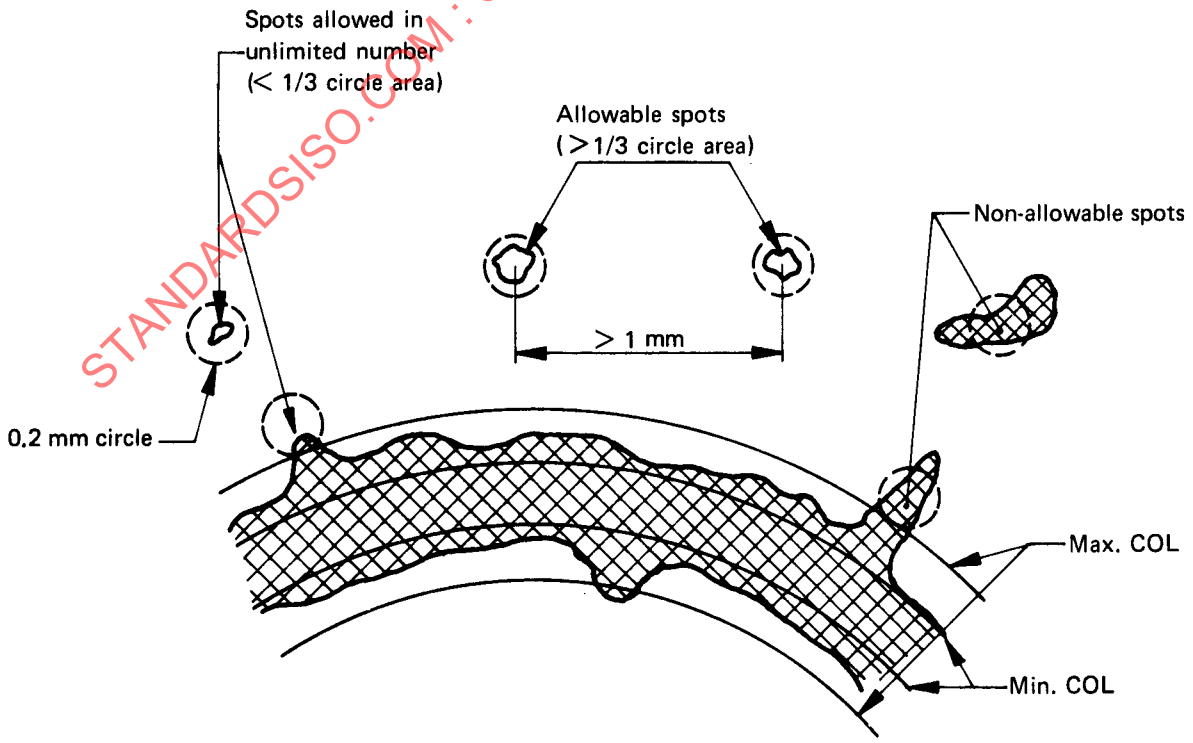


FIGURE 6

4.9 Voids

4.9.1 *Definition.* Voids are defined as areas inside the minimum COL which can be identified as being of lower density than the printed image. Voids can be measured either visually or in terms of PCS.

Both methods constitute an acceptable procedure for measurement of the printed image for most printing. However, where there is a substantial question of subjective visual judgment the instrumental method based on PCS must be used.

4.9.2 *PCS measurement of voids.* Sizes of voids depend on the PCS level at which they are measured. For a specific PCS level (see Appendix Y, clause A.2.7.1), a void is described as "allowable" if it satisfies the following conditions :

(1) The distance the PCS measuring aperture can be moved in a straight line, so as to give a PCS below the chosen level, nowhere exceeds 0.2 mm (0.008 in). The measuring aperture must remain at all times inside the minimum COL.

(2) The distance, centre to centre, of the void from the nearest other void, detected at the same PCS level, is at least 1.0 mm (0.040 in).

"PCS voids" is defined as the maximum PCS level at which all voids are allowable (see Figure 5).

4.9.3 *Visual measurement of voids.* A void is allowable if it satisfies the following conditions :

(1) It can be contained entirely within a circle of 0.2 mm (0.008 in) diameter, estimated visually.

(2) Its distance, centre to centre, from the nearest other void is at least 1.0 mm (0.040 in).

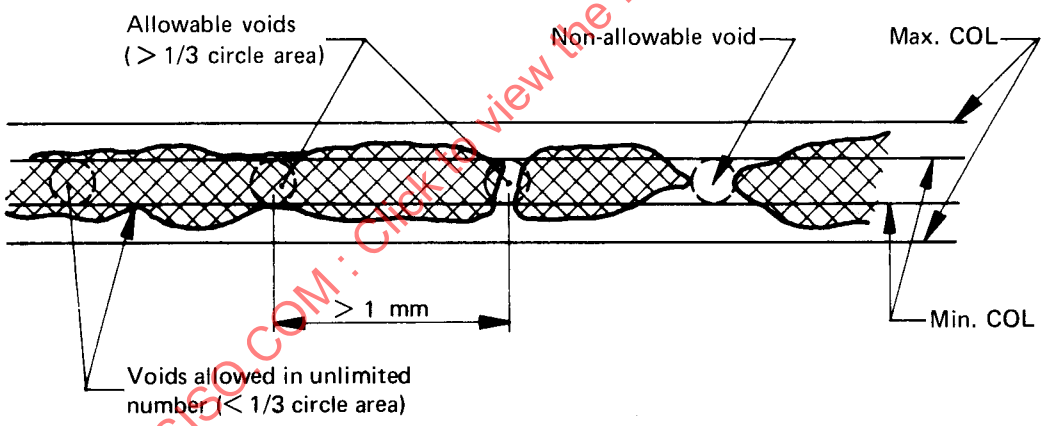


FIGURE 7

Small voids or groups of small voids which are contained in a circle of 0.2 mm (0.008 in) diameter are allowed in unlimited number if the total area of the void(s) contained in the circle of 0.2 mm (0.008 in) diameter is smaller than one-third of the area of the circle (see Figure 7).

4.10 Maximum PCS

The maximum PCS is the highest PCS level which is continuously exceeded for a scanning distance of 0.2 mm (0.008 in) within the maximum COL.

4.11 PCS of spots and voids

PCS spots and PCS voids as defined in clauses 4.8.2 and 4.9.2 must satisfy the following conditions :

(1) $\frac{\text{PCS voids}}{\text{PCS spots}} \geqslant 1.3$

(2) PCS voids $\geqslant 0.3$

It should be recognized that these are minimum values and that higher levels can generally be achieved, (see Appendix Y, clause Y.2.7).

4.12 PCS voids and maximum PCS

PCS voids and maximum PCS as defined in clauses 4.9.2 and 4.10 should preferably satisfy the following conditions :

$$\frac{\text{PCS max.}}{\text{PCS voids}} \leqslant 1.75$$

4.13 Average PCS within a character

In addition to the requirements stated in clauses 4.11 and 4.12 at least 80 % of the PCS within a minimum COL (that is, 80 % of the measurement made along the centreline) should preferably be greater than 0.4.

4.14 Character skew

The skew of a character is the rotational deviation of the printed image from its intended orientation relative to the document reference edge. Character skew must not exceed ± 3°.

5. CHARACTER POSITIONING

5.1 General

This section contains basic specifications relating to the position of characters on a document to accommodate general requirements of OCR devices.

It does not contain all the rules which may be necessary for a particular application, (see clauses 1.2 and 1.3).

5.2 Document reference edge

Some specifications in this section relate to document reference edges.

These can be horizontal and/or vertical edges, preferably the bottom and/or right-hand edges. Character alignment is relative to these reference edges.

5.3 Clear area

A clear area is defined as that region of a document reserved for the OCR characters and the clear space around these characters. The locations and dimensions of clear areas will be determined by the nature of individual applications and the requirements specified in this section.

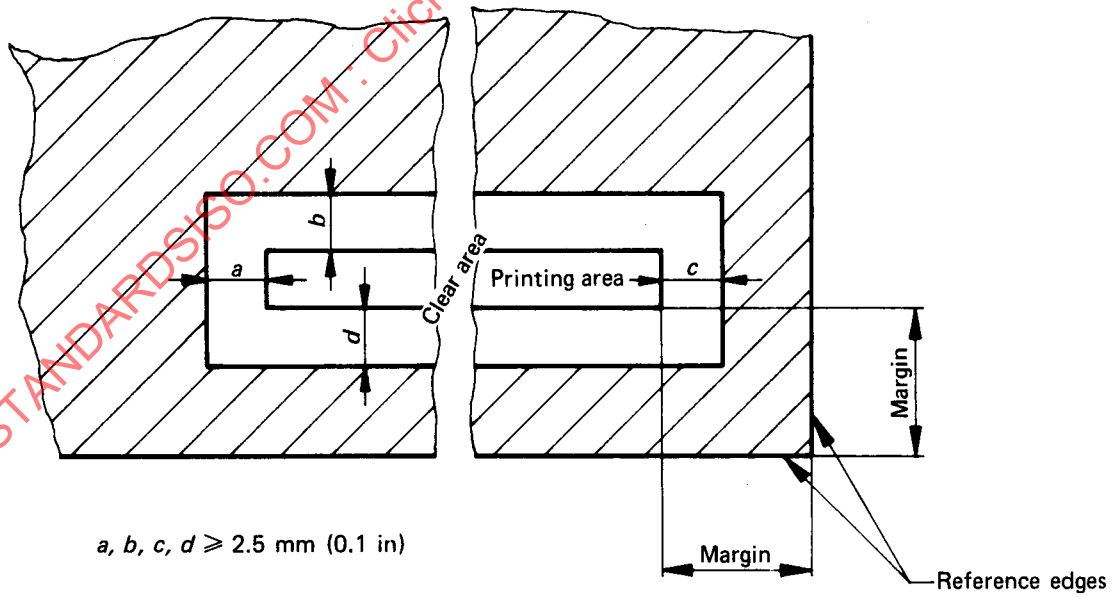


FIGURE 8

5.4 Printing area

A printing area is a rectangle inside the clear area, in which only OCR characters are to be printed. The sides of this rectangle must be parallel or perpendicular to a document reference edge (see Figure 8).

The distance between the corresponding boundaries of the printing area and the clear area should not be less than 2.5 mm (0.1 in).

5.5 Margin

The margin is the distance between the boundaries of the printing area and any paper edge, (see Figure 8).

A margin should preferably be not less than 6 mm (0.236 in). Where a specific application imposes a smaller value, the absolute minimum is 0.36 mm (0.014 in). In this case special care must be given to the compatibility of the print and the reading equipment.

5.6 Field

A field is a specified portion of the printing area that is limited to sets of one or more characters that may be treated as a unit of information. These character sets must be located in a single line of printing. A line could comprise several fields (see Appendix Y, clause Y.3.9).

5.7 Field boundary

A field boundary is defined as the smallest rectangle with sides parallel and perpendicular to a document reference edge, which contains all the boundaries of the component characters of the field.

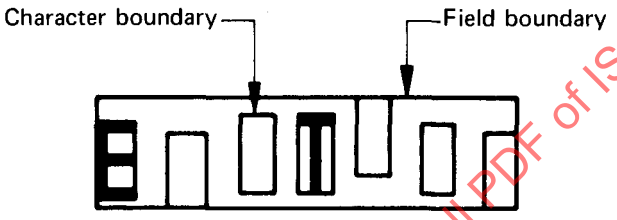


FIGURE 9

5.8 Line spacing

Line spacing is the vertical distance between the average horizontal centreline positions of all OCR characters printed on one line and that of all OCR characters printed on the next line (see Figure 10).

Nominal line spacing must be selected in such a way as to comply with the line separation tolerance (the parameters which influence line separation are line pitch specification, line skew, vertical misalignment, character height and strokewidth).

In any case the line spacing shall not be less than 4.0 mm (0.157 in).

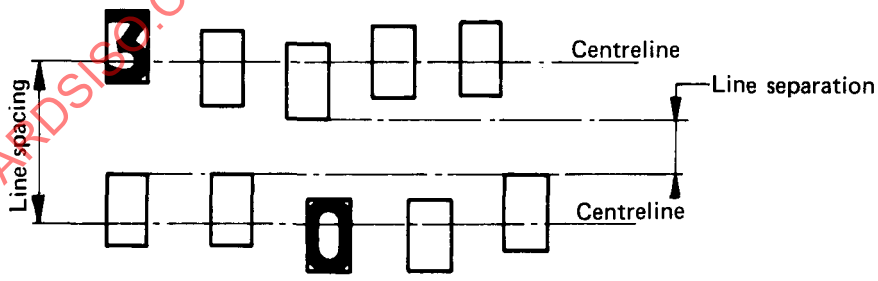


FIGURE 10

5.9 Line separation

Line separation is the vertical distance between the highest OCR character boundary (see clause 5.10) in a line and the lowest OCR character boundary in the line immediately above (see Figure 10).

The line separation should be not less than 2.5 mm (0.1 in). When closely spaced lines are necessary (e.g. for pages) a smaller separation may be inevitable, but the line separation should be maintained as large as possible by means of a reduction in vertical misalignment of the characters and by close conformity to the nominal strokewidth specification.

The minimum line separation shall not be less than the following values :

Size		I	II	III	IV
Minimum line separation	mm	0.64	1.0	1.5	2.0
	in	0.025	0.04	0.06	0.08

If the character sizes are intermixed, the line separation limitation for any pair of lines shall be that applicable to the largest character in the two lines.

5.10 Character boundary

The character boundary is defined as the rectangle with sides parallel and perpendicular to a document reference edge which is drawn tangential to the character outline and contains the character completely.

Skewed characters still have boundaries parallel or perpendicular to a document reference edge.

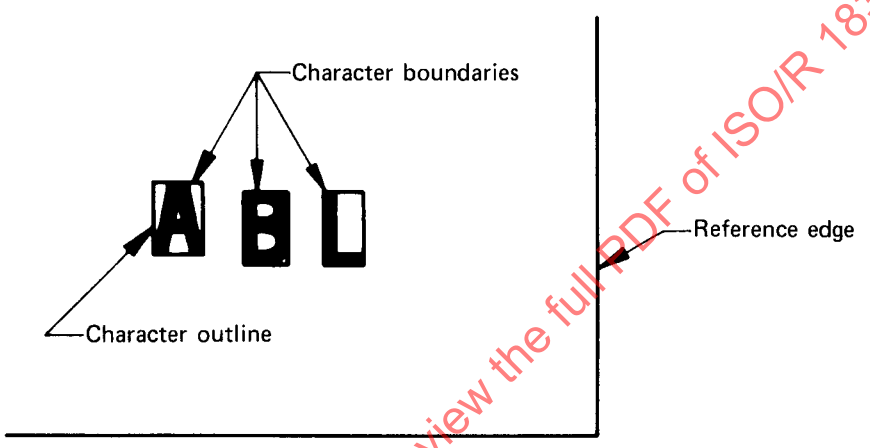


FIGURE 11

For the purpose of determining the boundary of the long vertical mark, only that portion of the long vertical mark will be considered which lies between the extension of the uppermost and lowermost horizontal boundaries of the adjacent character(s).

The character boundary is used to measure character and line separation and to determine field boundary.

5.11 Character reference lines

Character reference lines are used to determine the position of a character relative to some other character or to some reference edge.

5.11.1 *Character alignment reference line.* The character alignment reference line is the horizontal centreline or the lower edge of the character boundary (see Figure 13).

5.11.2 *Character spacing reference line.* The character spacing reference line is the vertical centreline of the character boundary.

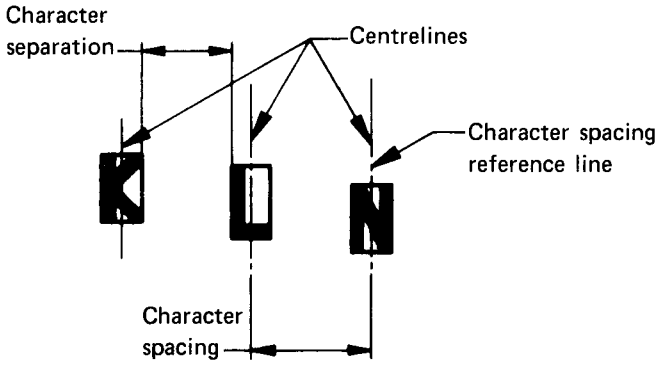


FIGURE 12

5.12 Character spacing

Character spacing (see Figure 12) is the horizontal distance between the character spacing reference lines of two adjacent characters (including the long vertical mark) corrected by the distance which would exist between the character spacing reference lines if the same two characters were superimposed in their nominal position. (This correction is derived from the nominal drawings and from the references used for the nominal alignment.)

For a character where the vertical reference line given in the nominal drawings coincides with its character spacing reference line, this correction does not apply.

Two characters are adjacent if the distance between their character spacing reference lines, corrected as mentioned above, is smaller than the following maximum values :

Size		I	II	III	IV
Maximum spacing	mm	4.57	4.57	4.57	6.60
	in	0.180	0.180	0.180	0.260

Character spacing of all characters shall not be less than the following specified minimum values :

Size		I	II	III	IV
Minimum spacing	mm	2.29	2.29	2.29	3.30
	in	0.090	0.090	0.090	0.130

Character spacing specifications will not be met when variable pitch printing is used (e.g. letterpress variable pitch typewriters; see Appendix Y, clause Y.3.7).

5.13 Character separation

Character separation is the horizontal distance between the adjacent boundaries of any OCR character(s) and/or the long vertical mark (see Figure 12). The character separation shall not be less than the nominal strokewidth as specified in clause 4.5.

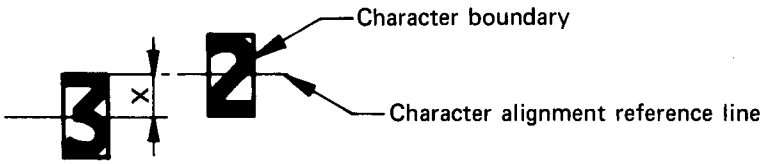
5.14 Character misalignment

Character misalignment is the vertical distance between the character alignment reference lines of two characters in the same line, corrected by the distance which would exist between the character alignment reference lines if the same two characters were printed in their nominal position. (This correction is derived from the nominal drawings and from the references used for the nominal alignment. However, it will not be needed, for instance, when the print is purely numeric or purely upper case alphabet. In other cases it should be determined whether a correction is necessary.) (See Appendix Y, clause Y.3.9.)

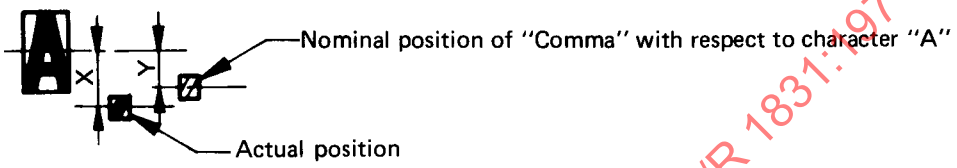
5.14.1 *Adjacent character misalignment.* Adjacent character misalignment is measured according to the above procedure between the character alignment reference lines of adjacent characters (see Figure 13). It shall not exceed the following values :

Size		I	II	III	IV
Max. adjacent character misalignment	mm	0.66	0.66	0.89	1.07
	in	0.026	0.026	0.035	0.042

This specification applies only to fields (see clause 5.6).



X = adjacent character misalignment



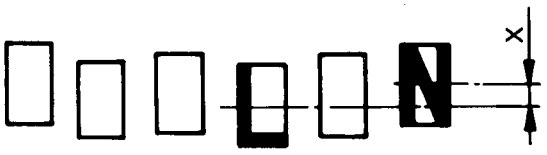
(X – Y) = Adjacent character misalignment

FIGURE 13

5.14.2 *Character misalignment in a field.* Character misalignment within a field is measured according to the above procedure between the alignment reference lines of any two characters in a field (see Figure 14). It shall not exceed the following values :

Size		I	II	III	IV
Max line character alignment	mm	1.32	1.32	1.78	2.14
	in	0.052	0.052	0.070	0.085

This specification applies only to fields (see clause 5.6).



X = character misalignment in field

FIGURE 14

5.14.3 *Long vertical mark alignment.* The long vertical mark must extend beyond the top and the bottom boundaries of any adjacent character.

A long vertical mark in one field should not extend nearer than 2.5 mm (0.1 in) to a field boundary in an adjacent line to which it does not apply.

APPENDIX Y

ADDITIONAL INFORMATION

Y.1 PAPER CHARACTERISTICS AND MEASUREMENTS

Y.1.1 Spectral properties

Y.1.1.1 *Significance of spectral properties for OCR documents.* An OCR scanner will usually be responsive to a restricted band of optical wavelengths. Typically, these scanners respond to the near ultra-violet, the blue-green and green or the near infra-red wavelengths.

Therefore, it is a fundamental requirement that the paper used for an OCR document be a good reflector in the wavelength ranges of the optical scanner response.

Y.1.1.2 *Colour.* It is strongly recommended that the paper for an OCR document be white. White paper is essentially non-selective to wavelengths of light within the range of interest for OCR scanners. Consequently if white paper is used no conflict of spectral properties will occur.

The specification excludes the use of most coloured papers, especially those with a definite and positive visual indication of colour.

If the saturation of the colour is slight, and the colour is essentially uniform throughout the OCR area on the documents, it is possible that they will comply with the specifications on average reflectance.

Y.1.1.3 *Notes on measurements.*

Y.1.1.3.1 MEANS OF REALIZING B 400. The light source specified in the body of this ISO Recommendation may be realized by a P 16 or Q type of cathode ray tube. The detector specified may be an S 4 or S 11 photodetector. The spectral response of the detector is extended to 500 nm in order to detect all the reflected energy as well as any energy resulting from fluorescent additives which may be present in the paper under evaluation

Y.1.1.3.2 MEANS OF REALIZING B 900. To implement the B 900 measurements the following components may be used :

- illumination source : incandescent lamp;
- sensor : silicon phototransducer;
- glass filter : a low frequency pass filter with cut-off at about 800 nm.

Y.1.1.3.3 ABSOLUTE WHITE REFERENCE. For most practical purposes magnesium carbonate ($MgCO_3$), which has a reflectance value very close to magnesium oxide (MgO), may be used instead of MgO for the 100 % reflectance reference without significant loss of accuracy.

Y.1.1.3.4 FLUORESCENT ADDITIVES. Fluorescent additives should be excluded in the manufacturing of paper for OCR use. However, it is realized that some contamination may be derived from previous operations. An effort should be made to minimize this contamination.

Y.1.2 Uniformity of paper reflectance

Y.1.2.1 *General.* Most scanner systems for OCR will examine in detail the area containing the printed image.

The reflected light from small areas on the paper of the order of 0.1 mm (0.004 in) diameter constitutes the input to a photodetector. The presence of ink is determined by a significant change in the reflectance of these areas relative to the paper.

Paper normally has a variation of reflectance, on this small area basis, because of the formation of the fibre structure and may have similar variations due to its surface characteristics, embossment of patterns or the printing of coloured patterns.

It is important that the magnitude of any such variations be significantly lower than the magnitude of differences between the paper and the printing.

Embossment of patterns or the printing of coloured patterns should be strongly avoided.

Y.1.2.2 *Measurement*

Y.1.2.2.1 REFLECTANCE OF THE BLACK BACKING. The backing of 0.5 % reflectance is most easily provided by an unlit cavity. If this is impracticable and if a solid with this low reflectance is not available, solids with a higher reflectance (not more than 3 %) may be used without serious loss of accuracy.

Y.1.2.2.2 PROCEDURES. The distribution of the reflectance values measured over equal areas of a sample of paper closely approximates to a normal curve. Such a normal distribution is defined by its mean (average paper reflectance measured with a black background) and its standard deviation σ (variation in paper reflectance).

If discrete measurements are made separated by at least 2 mm (0.008 in) they can be considered as being non-correlated. The number of observations required for reliable determination of σ is then of the order of 200.

If the observations are taken in one or more continuous scans, it is required that a total scanning length of at least 20 or 40 cm (8 or 16 in) be covered, for high or medium opacity papers respectively, in order to avoid the influence of interference between the reflectances of neighbouring points. This corresponds approximately to 200 non-overlapping points.

A procedure which avoids the calculation of the standard deviation and may be found more convenient in practice (still being sufficiently accurate) is as follows :

(1) Arrange the measurements obtained by one of the above scanning methods in a descending order of magnitude.

(2) Exclude the highest 0.5 % and the lowest 0.5 % of the values. Calculate the ratio R_{max}/R_{min} for the remaining values.

(3) This ratio should not exceed 1.2 for the high opacity class and 1.3 for medium opacity class.

For paper of "wild" formation with high variation in transparency the distribution of reflectance values may appreciably deviate from the normal distribution. In these cases the procedure which avoids the calculation of σ will be more satisfactory.

Y.1.2.3 *Recommendation.* The values given in this ISO Recommendation should be considered as lower limits. For reliable recognition more stringent limits are recommended and are obtainable with high quality paper.

Y.1.3 **Paper opacity**

Y.1.3.1 *Significance of paper opacity.* The opacity is indicative of the change in paper reflectivity on an OCR document due to the backing material at the time of scanning. If the document transport system of the OCR device is such that a known uniform reflective surface is provided at the time of scanning, a moderately opaque paper may be usable.

However, some systems scan the document while backed by other printed documents or have a transplant system that provides a non-uniform backing surface. For such cases a more opaque paper should be used, or a higher PCS value should be required for OCR information.

Y.1.3.2 *Recommendations.* The minimum opacity required for an OCR paper will be dependent upon the means of scanning and the application. In general, opacity is related to the basis weight of the paper; the higher the basis weight the greater the opacity. Consequently, there is a similar relationship between opacity and paper thickness, although the use of filler and coating materials has an effect.

In general, paper having opacity exceeding 85 % should be used. Papers of lower opacity should be used only if needed for the application and after considering the scanner optical system. Papers having opacity less than 65 % should not be used.

Many inks have the property of permeating the paper to a considerable depth. Applications requiring an OCR document to be printed on both sides may require a higher opacity or thicker paper to compensate for this effect.

Y.1.4 Paper gloss

Y.1.4.1 *Significance of gloss for OCR documents.* Gloss is the property of a surface responsible for a lustrous or mirror-like appearance. It is a phenomenon related to the specular reflection of incident light. The effect of gloss is to reflect more of the incident light in a specular manner, and to scatter less. It occurs at all angles of incidence and should not be confused with grazing angle specular reflection, which is often referred to as sheen.

Paper gloss is undesirable for OCR systems since it will change the effective brightness of the paper, thus affecting the print contrast signal.

Y.1.4.2 *Recommendations.* Paper for OCR documents should be restricted to the low gloss varieties. The use of coated or super-calendered papers or other papers with a glossy appearance should be avoided.

Y.1.5 Dirt in paper

Dirt in paper refers to the presence of relatively non-reflective foreign particles embedded in the sheet. Generally these particles are quite small and infrequent in good quality paper. The frequency of their distribution is significant.

The size and lack of reflectance of the particles may be such that they will be mistaken for inked areas by an OCR scanner.

Y.1.6 Mechanical properties of paper

Some mechanical properties of paper, such as tear resistance, bursting strength, folding resistance, etc., may be significant in OCR applications.

It is advisable that there be agreement on the specific papers intended to be used for such applications between users and manufacturers of OCR systems.

Y.2 CHARACTERISTICS OF THE PRINTED IMAGE

Y.2.1 General

This ISO Recommendation specifies the requirements for optimum reading system performance.

The specifications should be met by all print as far as is possible in the presence of the random effects which occur in any printing process.

The design of printers and the selection of supplies should ensure maximum compliance with the specifications. In any system the specifications may occasionally be exceeded, but the frequency with which this is allowed to occur should be carefully studied in the light of the reader performance required.

Y.2.2 Rules for the design of measuring gauges for OCR characters

The gauges are constructed by superimposing the minimum and maximum strokewidths, as specified in this ISO Recommendation, symmetrically about each point on the character centreline drawings.

The following rules also apply :

(1) "Internal angles" of maximum COL and "external angles" of minimum COL should be rounded with a radius of 0.1 mm (0.004 in).

(2) When the character centreline presents a sharp corner, the "external angle" of maximum COL and the "internal angle" of minimum COL should be identical to the angle of the character centreline.

NOTE

– "External angles" means angles which lie in the region where the angle determined by the character centreline is greater than 180°.

– "Internal angles" means angles which lie in the region where the angle determined by the character centreline is smaller than 180°.

– The centreline of the character may be included in the gauge to help in finding the best fit and in measuring centreline deviations.

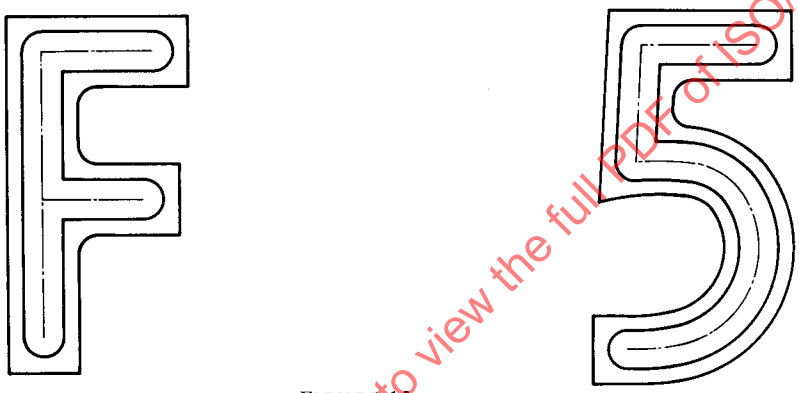


FIGURE 15

Y.2.3 Centreline deviations

The tolerance on centreline deviations is specified in this ISO Recommendation in order to limit the allowed deformation in the printed character caused by uneven printing and to give guidance on the choice of tolerances in the manufacture of type.

When this tolerance is met the printed edges will be symmetrically distributed along the centreline of the gauges aligned to give the "best fit".

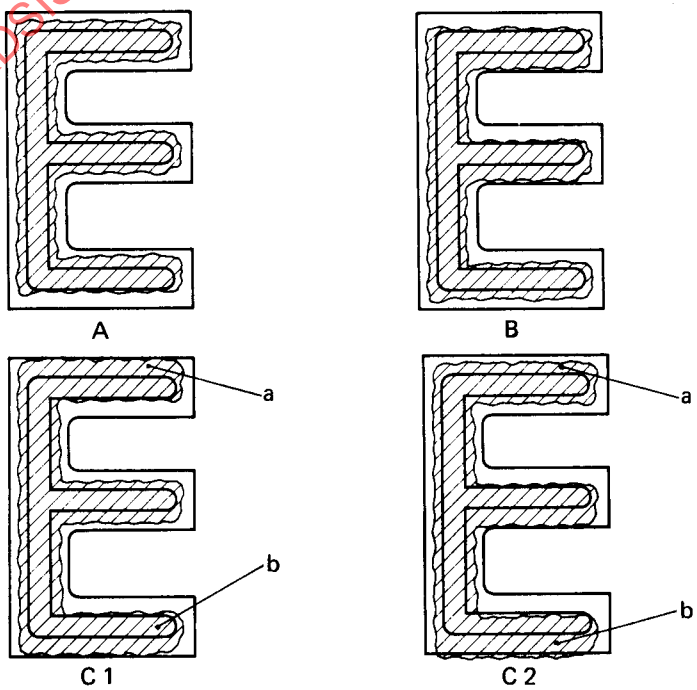


FIGURE 16

The print of examples A and B satisfies the specification. The print of examples C1 and C2 is out of specification because when the gauge is centred on the stroke "a" the stroke "b" will not be contained within maximum and minimum COL.

The tolerance may be checked by means of the gauge with the following procedure (see Figure 16) :

(1) When the gauge is centred on any stroke of the character, so that the edges are symmetrical to the corresponding centreline, all the rest of the character edges must lie between maximum and minimum COL.

(2) The above condition must be satisfied on centring the gauge on any stroke or portion of the character under examination.

Y.2.4 Strokewidth ranges

The variation in strokewidth from the nominal should be held to a minimum, since generally this could have a bearing on the reader performance.

Strokewidth range X requires a high quality printing process and careful control of maintenance and supplies. It cannot be met by some printers in common use for OCR. However, the tolerances which these printers normally produce do not necessarily extend to the full range Y. In such cases, printing performance should not be allowed to degrade beyond the normal level.

Y.2.5 PCS and visual measurements

In drawing up this ISO Recommendation it was recognized that the majority of print measurements will be made visually. At the same time an objective method of measurement, closely related to the way in which optical scanners operate, was required for critical evaluations.

The PCS specification fulfils the requirements for scanner-related measurements. Every effort has been made to ensure that the two specifications are equivalent and they are intended to be useful independently. Exact correlation is not possible, however, and some differences will arise (see clause Y.2.7).

Y.2.6 Spectral bands for PCS

For machine recognition of printed information it is necessary that a good contrast exist between the printed image and the paper. This contrast, expressed in PCS, is obtained when the paper has a good reflectance and the print is dense enough to provide a good absorption in the spectral range of interest.

Reading devices usually have a spectral response in the near ultra-violet, the visible or the near infra-red spectrum.

A printing ink provides good absorbance in one or more of these bands, depending on its composition. For example, black pigments tend to absorb light in all three bands, but dyes are more selective and usually yield the best absorption in the visible region.

Because of the diverse nature of printing equipment and OCR systems it is impossible to specify a single spectral range which contains the spectral responses of all reading devices and in which all printing inks would absorb sufficiently.

Which of the three specified spectral bands should be used, therefore, depends on the reading and printing devices in the application concerned. The following considerations apply :

(*a*) If the characteristics of all readers in the system are known, it is sufficient to choose the spectral band(s) appropriate to these readers.

(*b*) Printing which is required to satisfy the PCS specifications in the visible range imposes the least restriction upon the spectral characteristics of the printing inks.

(*c*) The only print which can meet the spectral requirements of all reading systems is that which conforms to the specification in all three bands. Print on white paper with ink of a high carbon black content will in general meet this requirement. This consideration also applies in applications where the reading systems to be used are not known when the application is to be established.

PCS measurements in the near ultra-violet region have to be implemented in the band B 425. However, it is realized that in some cases it might be convenient to use a band slightly shifted toward B 400. It must be noted that such measurements might give lower PCS values than those obtained using B 425.