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



For each individual country the only valid standard is the national standard of that country.

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.

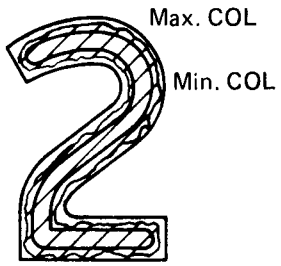



FIGURE 1 – Gauge in its "best fit" position

### 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

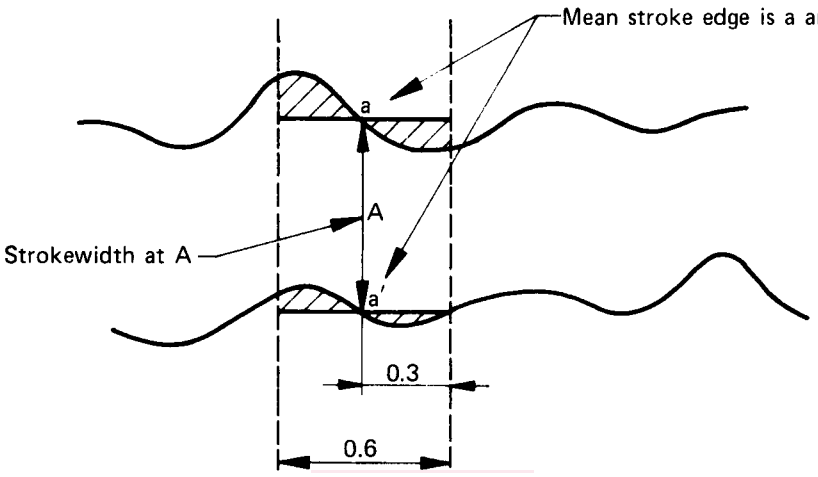


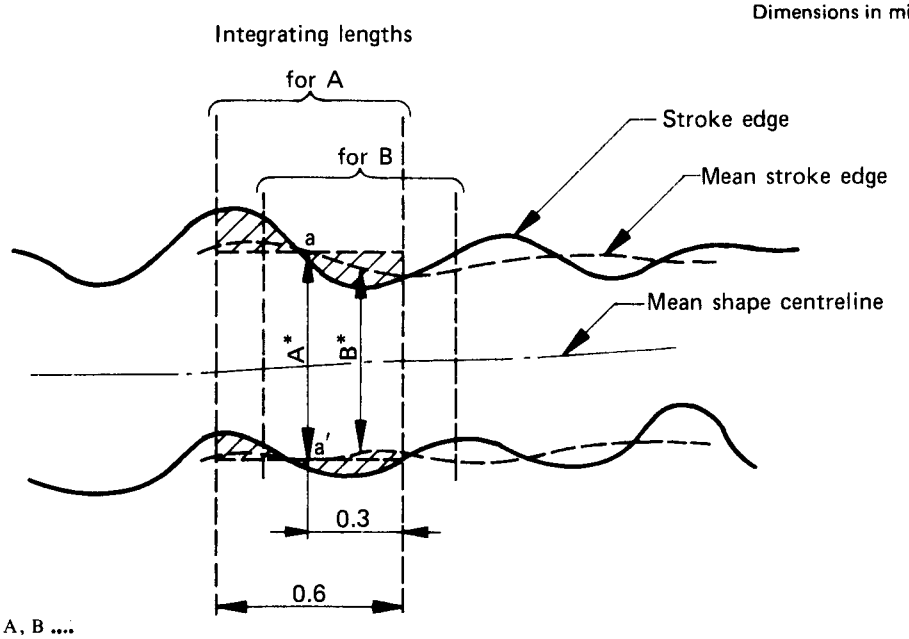

FIGURE 2

### 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



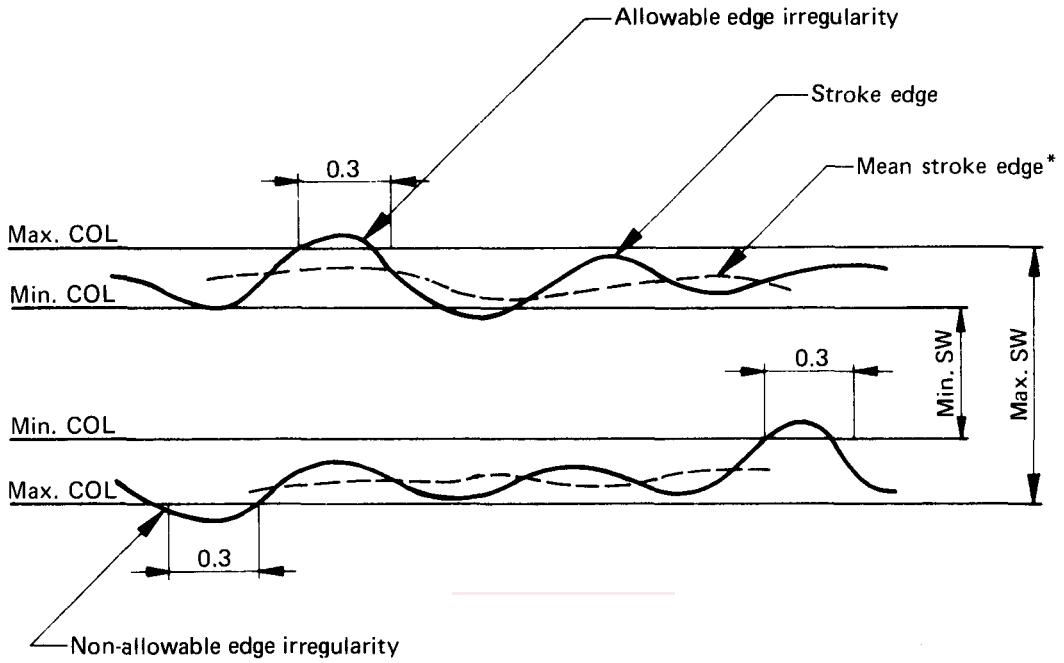

* Strokewidth is A, B ....

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 | + 0.008 |
| | | | | | - 0.11 | - 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 | + 0.008 |
| | | | | | - 0.11 | - 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 | + 0.008 5 |
| | | | | | - 0.14 | - 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).