International Standard 1831

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION•МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ•ORGANISATION INTERNATIONALE DE NORMALISATION

Printing specifications for optical character recognition

Spécifications d'impression des caractères pour reconnaissance optique

First edition — 1980-10-15

UDC 681.327.5.048 **Ref. No. ISO 1831-1980 (E)**

Descriptors : data processing, printing, optical characters, character recognition, optical recognition, printing papers, specifications, dimensions, legibility, measurement.

Price based on 42 pages

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards institutes (ISO member bodies). The work of developing International Standards is carried out through ISO technical committees. Every member body interested in a subject for which a technical committee has been set up has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work.

Draft International Standards adopted by the technical committees are circulated to the member bodies for approval before their acceptance as International Standards by the ISO Council.

International Standard ISO 1831 was developed by Technical Committee ISO/TC 97, *Computers and information processing*, and was circulated to the member bodies in January 1979.

It has been approved by the member bodies of the following countries :

Australia	Ireland	Romania
Belgium	Italy	South Africa, Rep. of
Czechoslovakia	Japan	Spain
Finland	Mexico	Sweden
France	Netherlands	Switzerland
Germany, F. R.	Poland	USSR

The member bodies of the following countries expressed disapproval of the document on technical grounds :

Canada
United Kingdom
USA

This International Standard cancels and replaces ISO Recommendation R 1831-1971, of which it constitutes a technical revision.

Contents

Annexes

Printing specifications for optical character recognition

0 Introduction

The purpose of this International Standard is to establish the basis for industry standards for paper and printing to be used in Optical Character Recognition (OCR) systems, particularly for document interchange, and to aid in the implementation and use of such systems.

It provides for the identification and measurement of, and establishes specifications for, the relevant parameters and gives guidance for their use.

0.1 Interpretation of the International Standard

A printing system is defined as a single unit comprising a printing machine, paper and inked ribbon (the latter only if required by the printing process). A printing system which produces printed material for OCR applications is called an OCR printing system.

The values in this International Standard shall apply to OCR printed material regardless of the printing system, font (OCR-A, OCR-B) and the specific application. The dimensional and optical characteristics of the printed image are given for three quality ranges.

Tolerance limits are specified for each parameter. These limits at least shall be achieved, but all parameters are expected to be kept well within them. If some of these parameters subject to variations of a statistical nature deviate from the specified limits, then the number and magnitude of these deviations can be reduced by using special precautions, such as a more accurate choice of the OCR printing system components, more frequent maintenance of the printing machine, a reduction of the printing speed, a shortening of the ribbon life, etc.

If the performance of an optical character recognition system is subject to variations of a statistical nature and if rejections or substitutions of characters within the tolerance limits occur then, again, the number and magnitude of these deviations can be reduced by using special precautions, such as a more frequent maintenance of the recognition system, etc.

0.2 Use of the International Standard

The measurement methods and the values of the parameters given in this standard are intended for use in OCR applications.

As a continuous, complete fulfilment of these values cannot be achieved because of the deviations of a statistical nature to which both printing and recognition systems are liable, some rejection and substitution of characters may occur. The number of rejections and substitutions which are allowed depends on the specific OCR application and shall be agreed upon, in statistical terms, by the user, the supplier(s) of the printing system and the supplier(s) of the recognition system.

In the guarantee of printing systems, the manufacturer of the printing system is given the right to specify the maintenance rate for the printing system and the supplies to be used (for example paper and ribbon).

In the guarantee of the recognition system, the supplier of the recognition system is given the right to specify the environmental conditions (temperature, humidity, illumination, maximum amount of mechanical vibrations and electromagnetic noise, etc.) and to establish the level of maintenance for the reader.

Statistical sampling plans by inspection of attributes can be used to check whether these guarantees are being observed, provided that these plans are coherent with those normally used in quality control.

Once a sampling plan has been agreed upon, the sample size (i.e. the number of characters or documents to be examined) is established by the plan.

To allow the printing system to be checked, the parameters of the printed material to be measured and the measurement methods are given in this International Standard.

When the recognition system is checked, only printed material meeting the specifications given in this International Standard shall be used, or — by agreement — representative samples of current material may be used. In the latter case the rejects must be evaluated according to their compliance with this International Standard.

0.3 Annexes

The annexes are not an integral part of this International Standard but give additional information.

1 Scope and field of application

This International Standard contains the basic definitions, measurement requirements, specifications and recommendations for OCR paper and print.

Three major parameters of a printed document for OCR media are covered. These are :

— the optical properties of the paper to be used;

— the optical and dimensional properties of the ink patterns forming OCR characters;

— the 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.

2 References

ISO 216, *Writing paper and certain classes of printed matter — Trimmed sizes — A and B series.*

ISO 1073/1, *Alphanumeric character sets for optical recognition — Part 1 : Character set OCR-A — Shapes and dimensions of the printed image.*

ISO 1073/2, *Alphanumeric character sets for optical recognition — Part 2 : Character set OCR-B — Shapes and dimensions of the printed image.*

ISO 2469, *Paper, board and pulps — Measurement of diffuse reflectance factor.*

ISO 2471, *Paper and board — Determination of opacity (paper backing) — Diffuse reflectance method.*

CIE Publication 15 (E 1.3.1) 1971 — *Colorimetry — Official recommendation.*

3 Spectral requirements

3.1 General

This clause defines spectral bands of interest for OCR applications.

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

3.2 Spectral bands

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, paper opacity and PCS measurement.

Table 1

Band	Peak nm	Bandwidth nm, 50 % level
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 shall be smooth curves without secondary peaks and with no major parts of the response curves 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.

4 Paper specifications for OCR

4.1 General

The papers to be used in OCR applications should be white (see annex A), have low gloss, and be of high opacity (see annex A). Factors causing variation in reflectance (such as dirt, uneven formation, watermarks and fluorescent additives) should be avoided.

In particular OCR applications, some mechanical properties of paper (such as stiffness, porosity, tear resistance and smoothness, etc.) may be important. For both optical and mechanical properties, agreement between users and manufacturers of OCR systems on the specific papers to be used is advisable.

4.2 Luminous reflectance factor R_o of paper

Reflectance measurements shall be carried out using a reflectometer as described in ISO 2469, or an instrument calibrated against such a reflectometer.

Reflectance measurements shall be referred to the perfect reflecting diffuser (100 % reflectance). However, in practice

barium sulphate ($BaSO_4$) may be used instead to give sufficient accuracy. In case of disagreement, the measurements shall be based on the perfect reflecting diffuser.

4.2.1 Definition of R_o

The luminous reflectance factor R_o is the reflectance factor obtained from a single sheet of paper using the black backing method, i.e. the sample being measured shall be backed with black having not more than 0,5 % reflectance.

The reflectance factor is the ratio, expressed as a percentage, of the radiation reflected by a body to that reflected by a perfect reflecting diffuser under the same conditions.

4.2.2 Measurement of R_o

R_o shall be measured using a method similar to that described in ISO 2471 but using the appropriate filters as described below.

4.2.3 Visual spectrum

R_o shall be greater than 60 % in the range from 425 to 500 nm and greater than 70 % in the range from 500 to 700 nm. For white, or slightly but uniformly coloured papers, it is normally sufficient to measure with the two following filters.

— B 425;

— CIE/Y filter, or any filter peaking between 530 nm and 570 nm and having a bandwidth not greater than 100 nm.

In case of doubt, measurements should be carried out throughout the visible spectrum using, for example, the filters B 425 to B 680 described in 3.2.

NOTE — If medium opacity paper (see 4.4.3.2) is used, the values for R_o shall be replaced by 50 % and 60 % respectively.

4.2.4 Near infra-red

When the near infra-red (IR) spectrum is of interest, R_o shall be not less than 70 % at 900 nm.

NOTE — If medium opacity paper (see 4.4.3.2) is used, the value for R_o shall be replaced by 60 %.

4.3 Dirt in paper

This refers to relatively non-reflecting foreign particles embedded in the sheet. Since the lack of reflectance and size of such particles may cause them to be mistaken for inked areas by an OCR scanner, it is important that both their frequency and size should be small.

Two methods of evaluating dirt in paper are described below. Method A enables a quick evaluation to be made whilst method B is suitable for a more detailed investigation.

For both methods the lighting conditions shall be according to CIE Publication 15.

4.3.1 Method A — Grid assay method

4.3.1.1 Equipment

This should consist of the following :

Grid

— A frame 1 m × 1 m (3.28 ft × 3.28 ft) divided into 100 squares by fine wire.

Working surface

— To accept paper and frame to allow viewing from a distance of about 0,5 m (1.64 ft).

Lighting

— The lighting should be a close approximation to the IEC recommended illuminant D 65. The recommended level of illumination is 750 to 1 500 lx.

Cleaner

— Soft brush or vacuum cleaner to remove loose dirt or dust from the sample surface.

Timer

— To indicate 0,5 or 1 min intervals.

Counter

— To tally the number of squares containing dirt.

4.3.1.2 Sampling and test area

Samples of a total of 6 m^2 (64.58 ft^2) shall represent the reel or stack of sheets. The reels shall be sampled at both ends with 6 × 1 m (3.28 ft) samples representing the full width for mill reels (sampling from the outer end of the preceding reel in manufacturing sequence if necessary). Sheet stacks shall be sampled at 6 positions with sufficient sheets to make up the total area.

4.3.1.3 Procedure

Lay out the sample with the topside uppermost.

Remove loose dirt and dust from the surface.

Place the grid over the sample.

Start the timer and scan all the squares in turn in 1 min. Record once only with the counter the number of squares seen to contain a dirt particle or particles.

Repeat the test on the remaining 5 m^2 (53.82 ft^2) and record as the number of squares containing dirt per 6 m^2 (64.58 ft^2). This number shall not exceed 200.

NOTE — For comparing results from different units, assessed samples should be exchanged for calibration between groups of observers. Observer-to-observer differences may exceed the variation due to change; observers can be selected by comparing assays and excluding observers giving significantly high or low variation. Observer comparisons should be made periodically.

4.3.2 Method B — Dirt count

The distribution of the dirt shall be established by a count of all light-absorbent surface particles above a certain size. A paper type fulfils the requirements of this International Standard when 20 samples have an arithmetic mean count of less than 250 dirt particles per m^2 (10.76 ft^2) with a diameter greater than 0,1 mm (0.004 in) each, and when 19 of these samples have, at the most, 25 particles per m^2 (10.76 ft^2) with diameters greater than 0,2 mm (0.008 in). The samples should preferably be equal to 1 m^2 (10.76 ft^2) but may not be less than 0,125 m^2 (1.345 ft^2), i.e. size A3, ISO 216. They shall be independent and provide a statistical representation of the full paper type to be evaluated.

4.4 Paper opacity

Opacity measurement shall be carried out using a reflectometer as described in ISO 2469, or an instrument calibrated against such a reflectometer.

4.4.1 Definition of paper opacity

Opacity (paper backing) is the ratio, expressed as a percentage, of the luminous reflectance factor R_o of a single sheet of the paper with a black backing to the intrinsic luminous reflectance factor R_∞ of the same sample of the paper. (This definition corresponds to that in ISO 2471.)

4.4.2 Measurement of paper opacity

The opacity shall be measured using the method described in ISO 2471. The filter used shall give, in conjunction with the optical characteristics of the basic instrument, an overall response equivalent to the spectral bands described in 3.2.

4.4.3 Classes of opacity

4.4.3.1 High opacity paper

High opacity paper shall have an opacity greater than 85 %.

4.4.3.2 Medium opacity paper

Medium opacity paper shall have an opacity greater than 70 % but less than 85 %.

4.5 Variation in reflectance of paper

Reflectance measurements performed with a very small aperture at a number of positions on the paper surface result in a variation of the measurements obtained.

These variations shall not exceed a given limit.

Due to their statistical nature, the limits for variation in paper reflectance are defined in terms of the allowable variation coefficient of the paper reflectance measured with an aperture of 0,2 mm (0.008 in) diameter.

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

For high opacity paper :

— standard deviation ⩽ 3,5 % of the mean reflectance (see 4.4.3.1);

For medium opacity paper :

— standard deviation ⩽ 5 % of the mean reflectance (see 4.4.3.2).

The specification on variation in paper reflectance shall 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, the three bands shall be used.

In addition, the ratio of the highest to the lowest value obtained by the measurements according to the above specification shall not exceed 1,2.

Detailed measurement procedures are laid down in annex A.

5 Characteristics of the printed image

5.1 General

In addition to the properties of the paper, the properties of the printed characters, i.e. the print quality, are critical in the recognition of characters. Characters to be read by optical recognition systems have to be of higher print quality than characters to be read by the human eye only. To achieve this higher print quality, appropriate inks, ribbons and printing machines shall be used and adequately operated and maintained.

Assessment of print quality shall include the examination of the geometry of the printed pattern (character shape) as well as the examination of the intensity of inking on the paper (print contrast). The characteristics of the ink (spectral response) are also of importance.

The characteristics described hereafter apply to the printed image, not to the printing device (for example type faces) with which the printed image is produced.

5.2 Print quality tolerance ranges

In general, the tolerances on print quality parameters in a successful OCR system will depend on the reader characteristics, on the required performance level and on the number of

characters in the reading repertoire considered. To accommodate these variations in capability of specific categories of printing and reading devices, three ranges of print quality are defined :

Print tolerance range X : tight tolerances

Print tolerance range Y : medium tolerances

Print tolerance range Z : wide tolerances

It should be noted that characters in range Z are reaching the limit of good quality print and are likely to give rise to an increased reject rate in many applications. Range Z characters can only be measured successfully by means of computer-aided methods (see 5.4.6).

5.3 Definition of character outline limits

The minimum and maximum character outline limits (COL) for a given character, in a specified font, character size and tolerance range, are the outlines of an ideal printed image of such a character with all the strokes having the respective stroke-width as specified in 5.3.1.

A COL gauge is a drawing on a transparent base of the two COLs and the centreline. Rules for the construction of COL gauges are given in 5.3.2 to 5.3.7.

5.3.1 Nominal strokewidth (see table 2)

For COL constructions the following nominal strokewidths and tolerances apply.

The heights indicated are exact for OCR-A. For OCR-B they are indicative; exact values shall be measured from the OCR-B drawings in ISO 1073.

For OCR-B, the nominal strokewidth of the small letters and of the characters #, %, @ is 0,31 mm (0.012 in), for size I and 0,44 mm (0.017 in) for size IV.

Table 2 — Nominal strokewidth

Size	Height		Nominal strokewidth		Tolerances ±			
					Range X		Ranges Y, Z	
	mm	in	mm	in	mm	in	mm	in
I	2,40	0.094	0,35	0.014	0,08	0.003	0,15	0.006
III	3,20	0.126	0,38	0.015	0,08	0.003	0,18	0.007
IV OCR-A	3,80	0.150	0,51	0.020	0,13	0.005	0,25	0.010
IV OCR-B	3,60	0.142	0,50					

5.3.2 Construction of the COL gauges

For a given character size and tolerance range the minimum COL is the geometric envelope of a circle of diameter equal to the minimum strokewidth centred on and moved along the character centreline. Likewise the maximum COL is the geometric envelope of a circle with a diameter equal to the maximum strokewidth centred on and moved along the character centreline.

Deviating from the general rules, the following rules apply to free ends of strokes and to corners of the stroke centreline of the gauges. These rules refer to "external" and "internal" corners which are defined as follows :

An external corner is a corner where the angle defined by the strokes of the centreline is greater than 180° (see figure 1).

An internal corner is a corner where the angle defined by the strokes of the centreline is smaller than 180° (see figure 1).

5.3.3 Fairing radii

The following fairing radii shall be used as indicated in 5.3.4 and 5.3.5. The same fairing radii are used for the construction of OCR-A and OCR-B COL gauges.

Table 3

Size	Fairing radius, minimum COL R_1		Fairing radius, maximum COL R_2	
	mm	in	mm	in
I	0,10	0.004	0,10	0.004
III	0,10	0.004	0,13	0.005
IV	0,13	0.005	0,20	0.008

5.3.4 Special rules for minimum COL

When the minimum COL presents an internal corner with a radius equal to or less than R_1 (see 5.3.3), it shall be drawn with a sharp corner defined by the tangents to the envelope at the point where the radius changes from greater to equal to or smaller than R_1 (see figure 2).

5.3.5 Special rules for maximum COL

5.3.5.1 Internal corner

When the maximum COL presents a sharp internal corner or a radius smaller than R_2 (see 5.3.3), a fairing radius equal to R_2 shall be used (see figure 2).

5.3.5.2 External corner

When the centreline has a sharp corner, the external corner of the maximum COL shall be drawn as a sharp corner also (see figure 2). An exception to this rule applies if the stroke centreline has a corner with an angle of more than 305°. In this case, the external corner of the maximum COL shall be drawn as a tangent to the envelope perpendicular to the bisector of the corner defined by the stroke centreline (see figure 3).

5.3.5.3 Free stroke ends

At free stroke ends, the maximum COL shall be squared off by drawing the tangent to the envelope parallel and perpendicular to the corresponding free end of the character stroke centreline (see figure 2).

5.3.6 Letterpress font

The letterpress font characters of OCR-B may be checked with the same gauges, constructed according to the rules stated above, in range X, size I. Attention shall be given to the following special features :

5.3.6.1 The nominal strokewidth of the letterpress font is not constant, but may deviate from the nominal value of the constant strokewidth font in range X. These deviations are 5 % to 10 % of the nominal value and can be neglected.

5.3.6.2 The nominal stroke outlines of some characters end with sharp corners of considerably less than 90°. At these corners, the stroke edges may extend outside maximum COL and inside minimum COL. These extensions are allowed if they are not obviously due to voids or spots. The latter are subject to the relevant specifications. However, there is no specific set of gauges for the letterpress font.

5.3.7 Additional rules for the construction of COL gauges for range Z

As mentioned in 5.2, characters in range Z can only be measured reliably by means of a computer-aided method (see 5.4.6). In this case special COL gauges shall be used.

Printed images that do not fulfil the shape requirements as defined for range X and range Y may be recognized by OCR machines as deviations from the requirements given for range Y, provided that these deviations are within certain limits and that the character repertoire is restricted to numeric sub-sets. The deviations most commonly known in practice are asymmetrical violations of minimum COL on one side of the character (at the top or at the bottom or on the right side or on the left side) called cut-off. Such deviations may happen for example with high speed printers (see figure 4).

The limit for the allowed cut-off shall be given by cut-off limit lines (see figure 5). The cut-off limit lines define a rectangle which is of equal size for all characters for a given font and for a given size. The dimensions of this rectangle shall be given by the horizontal and vertical dimensions of the largest character measured along the character-centreline.

The dimensions are given below of the different fonts and sizes.

Table 4

Font	Size	Height		Width	
		mm	in	mm	in
A, B	I	2,40	0.094	1,40	0.055
A, B	III	3,20	0.126	1,52	0.060
A	IV	3,80	0.150	2,04	0.080
B	IV	3,60	0.142	2,10	0.83

NOTE — Characters with the minimum COL's within the rectangle defined above for the cut-off limit lines shall have no cut-off.

The horizontal position of the rectangle shall be centred on the vertical centreline of the characters of font A, and centred on the vertical reference line of the characters of font B.

The vertical position shall be defined by the distance d_v between the base line of the rectangle and the horizontal character reference line (see figure 5). The dimensions for distance d_v are shown below.

Table 5

Font	Size	Distance d_v	
		mm	in
A	I	0,00	0.00
	III	0,00	0.00
	IV	0,00	0.00
B	I	0,13	0.005
	III	0,18	0.007
	IV	0,20	0.008

In the measuring gauge, the cut-off limit lines for each character shall be defined only inside maximum COL. Examples are shown in figure 6.

For those stroke elements that are affected by cut-off, a cut-off centreline is defined as follows :

The cut-off centreline is the geometrical locus of all centres of circles that can be drawn between the cut-off limit line and the internal line of the non-violated minimum COL. On the intersection between the cut-off limit line and the minimum COL of a gauge stroke-element, the cut-off centreline must fit into the gauge centreline.

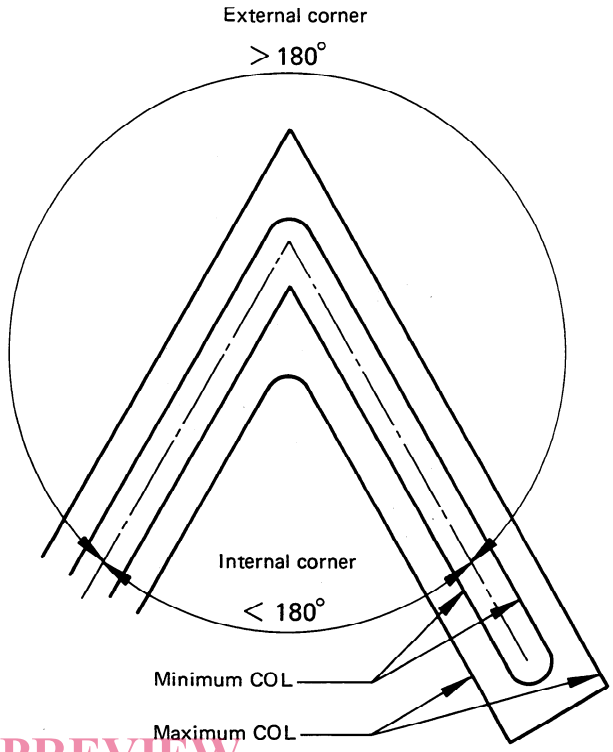

Figure 1 — Internal and external corner of stroke elements

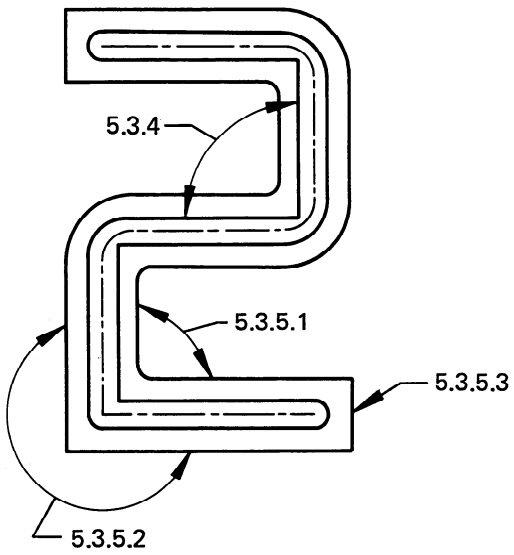

Figure 2 — Special situations at minimum and maximum COL

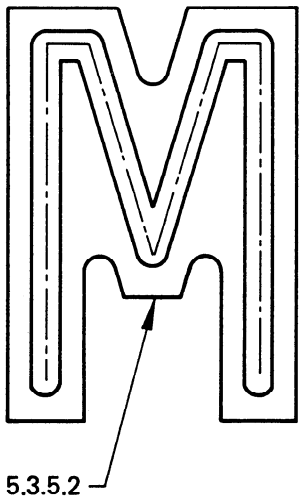

Figure 3 — Special corner at maximum COL

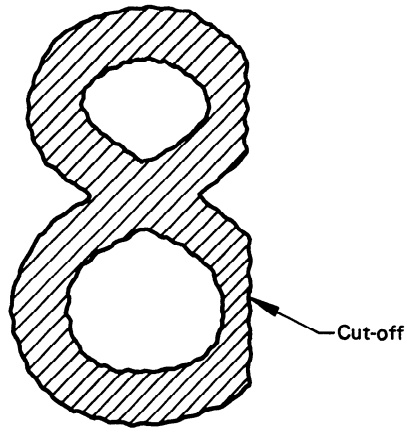

Figure 4 — Cut-off character

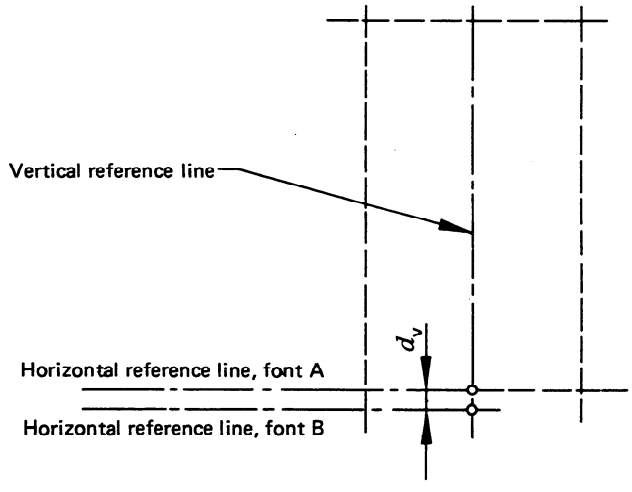

Figure 5 — Cut-off limit lines

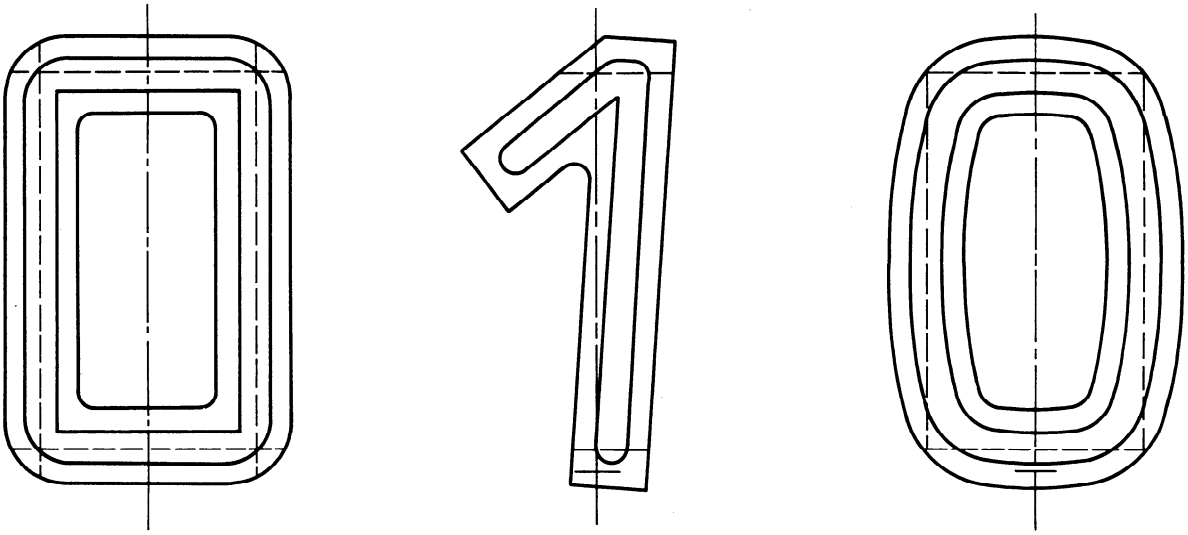

Figure 6 — Examples of gauges with cut-off limit lines

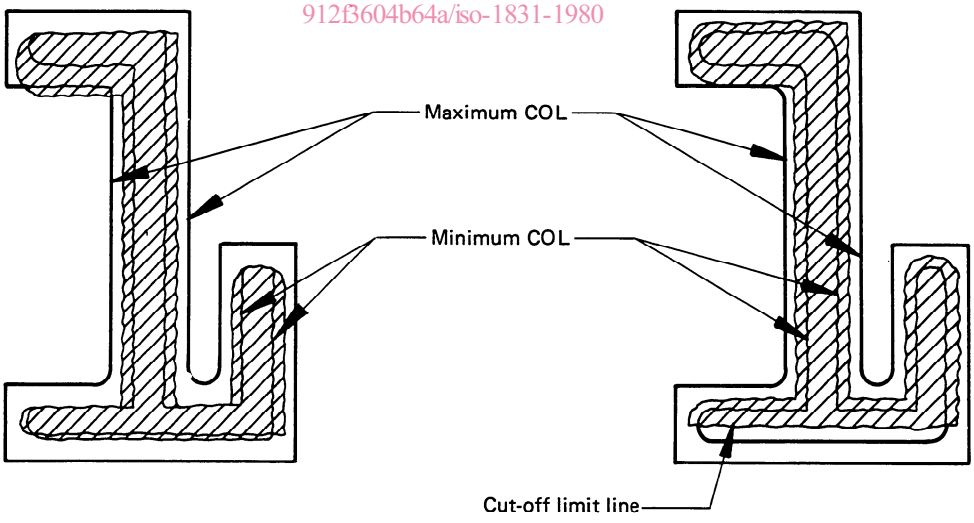

Figure 7 a) — Adjustment of a character without consideration of the cut-off limit line

Figure 7 b) — Adjustment of the same character with consideration of the cut-off limit line