ECMA

EUROPEAN COMPUTER MANUFACTURERS ASSOCIATION

STANDARD ECMA-15

for

PRINTING SPECIFICATIONS FOR OPTICAL CHARACTER RECOGNITION

2nd Edition — August 1975

Free copies of this ECMA standard are available from
ECMA European Computer Manufacturers Association
114 Rue du Rhône — 1204 Geneva (Switzerland)

ECMA

EUROPEAN COMPUTER MANUFACTURERS ASSOCIATION

STANDARD ECMA-15

for

PRINTING SPECIFICATIONS FOR OPTICAL CHARACTER RECOGNITION

2nd Edition — August 1975

BRIEF HISTORY

The first issue of Standard ECMA-15 dated May 1968 represented the international agreement reached within ISO/TC97/SC3 after years of work within this Committee. It corresponded to ISO/R 1831 1971. When the latter was published, it was recognized that at least parts of it should be reviewed. Unfortunately, international agreement within SC3 is this time far from being reached. It can be expected that no ISO standard will be achieved before a certain number of years.

In view of this situation, TC4 having developed a new simplified method for measuring printed image parameters and having checked this method by means of tests carried out in different Companies located in different countries, have decided to fill the present vacuum by issuing the present 2nd edition of ECMA-15. It is felt that there is an urgent need by OCR users to have at their disposal a workable document clearly defining the interface between printing and reading operations. The proposed method is very suitable for visual inspection of printed characters and is, by its very nature, relatively easy to implement by instrumental inspection as it requires inspection by discrete steps only. Suitable devices for PCS measurements exist and are commercially available.

This 2nd edition has been accepted by the ECMA General Assembly on June 19, 1975.

This issue supersedes the issue dated May 1968.

TABLE OF CONTENT

Table of Content (cont'd)

ANNEX

1. INTRODUCTION

1.1 PURPOSE

The purpose of this standard is to establish the basis for industry standards for paper and printing to be used in Optical Character Recognition (OCR) systems, 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.

1.2 SCOPE

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

It applies to the Standards ECMA-8 (Nominal Character Dimensions of the Numeric OCR-A Font) and ECMA-11 (Alphanumeric Character Set OCR-B for Optical Recognition). Three major parameters of a printed document for OCR media are covered. These are :

i) The optical properties of the paper to be used.

ii) The optical and dimensional properties of the ink patterns forming OCR characters.

iii) The basic requirement 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.

The positioning of OCR characters on documents is specified in the following ECMA Standards:

ECMA 18 - Printing line position on single line documents

ECMA 21 - Character positioning on OCR journal tape

The mechanical characteristics of the paper to be used for OCR application are given in the ECMA document "Recommended OCR paper specifications" of March 1970.

1.3 INTERPRETATION OF THE STANDARD

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

The figures in this standard apply to the OCR printed material, independently from the printing system, from the OCR font (OCR-A, OCR-B) and regardless of the specific application. The dimensional and optical characteristics of the printed image are given for two quality ranges.

Tolerance limits are specified for each parameter. These limits must at least be achieved, but all parameters are expected to be kept well within them. However some of these parameters are subject to variations of a statistical nature and deviations from the specified limits may occur. 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.

Likewise the performance of an optical character recognition system may be subject to variations of a statistical nature and rejections or substitutions of characters within the tolerance limits may occur. 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.

In both cases, these precautions are generally liable to increase the total cost of the OCR application and shall therefore be used only, and as far as, they are required. In other terms the degree to which the above mentioned deviations are allowed to occur shall be indicated for each specific OCR application.

1.4 USE OF THE STANDARD

The values of the parameters and the measurement methods given in this standard are recommended for use in OCR applications. Further work is in progress to cover other print quality ranges than those mentioned in this Standard. Computer aided measurement methods will also be investigated.

Since OCR is used in a broad range of applications, each user and each manufacturer must decide which of the parameters defined in the Standard are necessary for the specification of their system. It is recognized that some of the parameter values and tolerances might be too loose or too stringent for a given application.

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

Such an agreement shall establish that:

- the supplier of the printing system shall guarantee that a given "lot tolerance percent defective" LTPD (i.e. a given percentage of characters out of the relevant tolerance range) is liable to occur with a probability of less than a given value α;

- the supplier of the recognition system shall guarantee that a given percentage of characters within the relevant tolerance range is liable to be rejected with a probability of less than β and a given (obviously lower) percentage of characters within the relevant tolerance range is liable to be substituted with a probability of less than β.

In the guarantee of printing systems, the manufacturer of the printing system is given the right to specify the maintenance rate for the printing machine and the supplies to be used (e.g. 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 maintenance rate for the reader.

Statistical sampling plans by attributes can be used in order to check if these guarantees are observed; provided they are coherent with the sampling plans normally used in all other cases of 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; the elements of the sample must be selected by applying the following rules:

- all characters of the relevant repertoire shall be considered in the same proportion as they are expected to be present in the application;
- the characters (or documents) shall be selected at regular intervals during the ribbon life span (where applicable);
- the characters (or documents) shall be selected at regular intervals between two scheduled maintenances of the printing system;
- no adjustment of the printing system shall be made during the test, unless this adjustment is explicitly defined in the user's manual of the printing system.

When the printing system is checked, the parameters of the printed material to be measured, as well as the measurement methods, are given in this standard.

When the recognition system is checked, only printed material meeting the specifications in this standard shall be used. Tests shall be carried out at regular intervals between two scheduled maintenances of the recognition system; no other adjustments are allowed, unless explicitly defined in the user's manual of the recognition system.

2. SPECTRAL REQUIREMENTS

2.1 GENERAL

This Section defines 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

In this section 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 sections Paper Reflectance (3.2), Paper Opacity (3.4) and PCS Measurement (4.5.7.2).

BAND	PEAK in nm	BANDWIDTH in 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 must be smooth curves without secondary peaks and with no major parts of the responses 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.

3. PAPER SPECIFICATIONS FOR OCR

3.1 GENERAL

The papers to be used in OCR applications should be white, (see A.1.1.2) have low gloss, and be of good opacity (see A.1.3). Factors causing variation in reflectance (such as dirt, uneven formation, watermarks, fluorescent additives) should be avoided.

In particular OCR applications, some mechanical properties of paper (such as tear resistance, bursting strength, folding resistance, etc.) may be important. In these cases, agreement between users and manufacturers of OCR systems on the specific papers to be used is advisable. For document and journal interchange applications, this section should be used in conjunction with the "Recommended OCR Paper Specifications" published in March 1970.

3.2 PAPER REFLECTANCE

Reflectance measurements shall be carried out using a reflectometer as described in ISO 2469, or an instrument in calibration with 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 with a sufficient accuracy. In case of disagreement, the measurements shall be based on the perfect reflecting diffuser.

3.2.1 DEFINITION OF PAPER REFLECTANCE

The intrinsic reflectance factor of paper is the reflectance factor of a layer or pad of paper thick enough to be opaque.

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.

3.2.2 MEASUREMENT OF PAPER REFLECTANCE

The reflectance shall be measured using a method similar to that described in ISO 2470 but using the appropriate filters as described below.

3.2.3 VISUAL SPECTRUM

The intrinsic reflectance factor of the paper shall be not less than 60% in the range from 425 nm to 500 nm and not less than 70% in the range from 500 nm to 700 nm. For white, or slightly but uniformly coloured papers, it is normally sufficient to measure the intrinsic reflectance factor with the two following filters.

B 425

CIE/Y filter, or any filter peaking between 530 nm and 570 nm and having a band width 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 and B 680 described in Section 2.2.

3.2.4 NEAR INFRA RED

When the near infra red (IR) spectrum is of interest, the intrinsic reflectance value shall be not less than 70% at 900 nm. Since white and slightly coloured papers which meet the previous

specifications will usually meet this requirement, reflectance measurements at this wavelength are usually not necessary.

3.2.5 VARIATION OF PAPER REFLECTANCE

Variation of paper reflectance is the standard deviation of reflectance when measured using the black backing method. These measurements must be made over well separated circular areas with an instrument having an aperture of 0,2 mm diameter.

Two classes of variations of 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

This refers to relatively non-reflecting foreign particles embedded in the sheet. Since the lack of reflectance and size of such particules 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.

In the grid assay method described in A 1.4, the frequency of visible dirt is assessed in known conditions. For OCR applications, the number of squares found to contain dirt should not exceed 200 per 6 m^2.

3.4 PAPER OPACITY

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

3.4.1 DEFINITION OF PAPER OPACITY

Opacity (paper backing) is the ratio, expressed as a percentage, of the luminous reflectance factor of a single sheet of the paper with a black backing to the intrinsic luminous reflectance factor

of the same sample of the paper. (This definition corresponds to that in ISO Recommendation R 1912 - Paper Vocabulary).

3.4.2 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 CIE tristimulus value Y.

3.4.3 CLASSES OF OPACITY

3.4.3.1 HIGH OPACITY PAPER

High opacity paper has an opacity of not less than 85%.

3.4.3.2 MEDIUM OPACITY PAPER

Medium opacity paper has an opacity of not less than 70% but less than 85%.

4. CHARACTERISTICS OF THE PRINTED IMAGE

4.1 GENERAL

This section contains specifications and quality control criteria pertaining to individual OCR characters, 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 the latter 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. In order to achieve "good" print quality as suitable for OCR, it should be understood that in comparison with non-OCR applications special precautions are required, including adjustments, maintenance, more frequent changing of ribbons on some printing devices, and judicious selection of supplies. However, in terms of measurable quality, printed matters of "good" quality are those for which the requirements of the Standard are met.

The different type fonts, repertoires and sizes define a magnitude of different characters. In order to be able to measure the print quality of all these characters, visual and instrumented measuring methods are described.

The visual measurements can be conducted simply and fast with the aid of gauges. They are, therefore, given wide room in this Standard. It must be understood, however, that some parameters - mainly the contrast - can only roughly, sometimes unsatisfactorily and always only subjectively be estimated visually.

Therefore, in cases of conflict, the instrumented measurements should be conducted; these methods, although time consuming if all parameters are considered, yield objective and accurate results.

4.2 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 accomodate these variations in capability of specific categories of printing and reading devices, two ranges of print quality are defined:

Range X - high quality,

Range Y - medium quality.

All figures related to these print quality ranges are considered consolidated by experience.

4.3 DEFINITIONS OF PRINT CONTRAST

Print quality parameters fall into two basic groups, namely dimensional properties and optical propteries of the printed image. Measurement of the optical properties requires the definition of Print Contrast Signal (PCS). Measurement of the dimensional properties requires the definition of the Character Outline Limit (COL), see 4.4.

4.3.1 DEFINITION OF THE PRINT CONTRAST SIGNAL

Print Contrast Signal (PCS) is an expression which relates the contrast between any selected point and the background paper.

The PCS value of a point p is defined by:

$$PCSp = \frac{Rw - Rp}{Rw}$$

where:

Rw = the maximum reflectance fount within the area of interest to which the PCS of point p is referenced. In measuring printed images, the area of interest should be a rectangle of approximately twice the nominal character height by twice the nominal character width and centred on the character being measured (see Fig. 1).

Rp = the reflectance from a small measurement area centred on point p.

The reflectances Rw and Rp are measured within an area of 0,2 mm diameter if circular or 0,15 mm side if square.

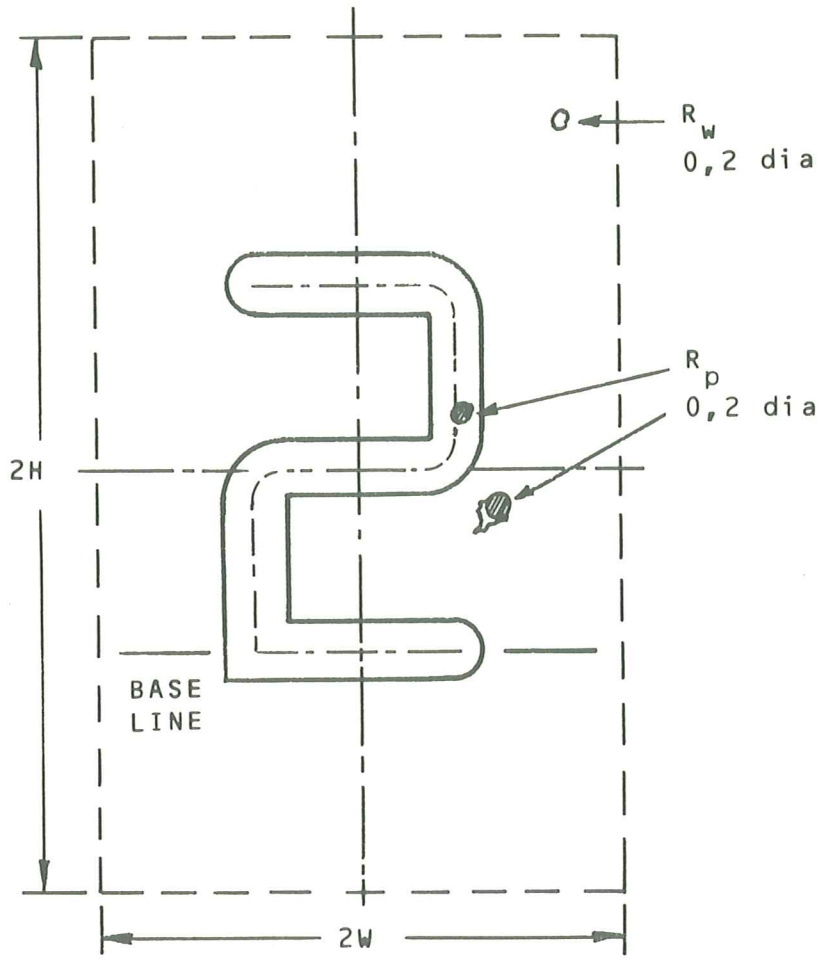


Fig. 1

These reflectance specifications deal only with diffuse reflectance and the reflected light used for measurement shall exclude specularly reflected light. Although normally reflectance values are referred to Ba SO_4 as the 100 % value, this is not necessary in the determination of PCS. The value of PCS is dependent only on the relative reflectance values of Rw and Rp. The reflectance measurements shall be made using the black backing method.

4.3.2 SPECTRAL REQUIREMENTS FOR MEASUREMENT OF PCS

The PCS of any point of a printed image is highly dependent on the spectral properties of the ink used to create the printed image. (See also A.2.4).

4.4 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 minimum and maximum strokewidth respectively, as specified in Table 2 (clause 4.5.8).

When constructed on a transparent base for visual use, the COL drawing is called a COL Measuring Gauge.

As to the construction of COL for each character, the following applies, (see also Fig. 2 to 5).

4.4.1 GENERAL RULE

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 moving along the character centreline. Likewise, the maximum COL is the geometric envelope of a circle of diameter equal to the maximum strokewidth and moved along the character centreline.

The following rules apply to free ends of strokes, corners in the centreline, and to characters of the letterpress font. These rules refer to "external" and "internal" corners.

An "external" corner is a corner which lies in the region where the angle defined by the strokes of the centreline is greater than 180° (see angle a in Fig. 3).

An "internal" corner is a corner which lies in the region where the angle defined by the strokes of the centreline is smaller than 180° (see angle b in Fig. 3).

4.4.2 MINIMUM COL

When the minimum COL presents a sharp internal corner or a radius equal to or less than R1, 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 R1 (see Fig. 2).

4.4.3 MAXIMUM COL

4.4.3.1 When the maximum COL presents a sharp internal corner or a radius of less than R2 (see Table 1), a fairing radius equal to R2 shall be used (see Fig. 2).

4.4.3.2 When two segments of the maximum COL are too close to eachother to use the fairing radius R2 (as may be encountered in parallel strokes joining into a cul-de-sac), then the largest possible fairing radius shall be used (see Fig. 3).

4.4.3.3 When the character centreline has a sharp corner, the external corner of maximum COL shall be drawn with an angle identical to that of the character centreline (see Fig. 2).

4.4.3.4 As an exception to Rule 4.4.3.3, when the character centreline has an internal sharp corner with an angle of less than 55^{o} , the corresponing external sharp corner of the maximum COL shall be squared off, by drawing the tangent to the envelope perpendicular to the bisector of such an angle (see Fig. 4).

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

4.4.4 LETTERPRESS FONT

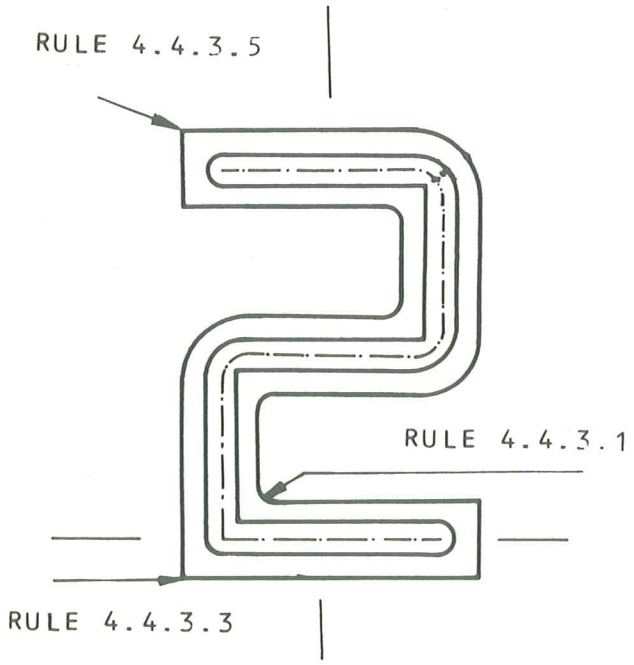
The letterpress font characters of OCR-B may be checked with the same gauges, constructed according to the rules stated before, in Range X, Size I. Attention should be given to the following special features:

4.4.4.1 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.

4.4.4.2 The nominal stroke outlines of some characters end with sharp corners of considerably less than 90^{o}. 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 (see Fig. 5). The latter are subject to the relevant specifications in sections 4.5.7 and 4.5.10.

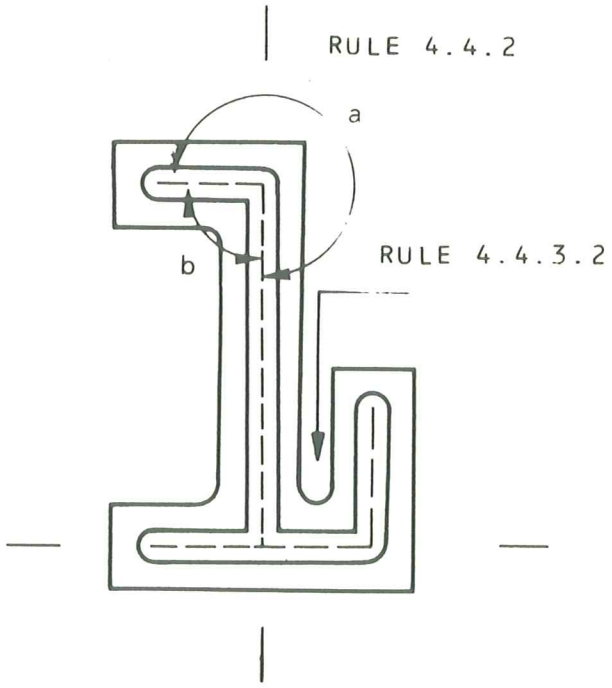
Table 1

OCR-A and OCR-B	R1 mm	R2 mm
Size I	0,10	0,10
Size III	0,10	0,13
Size IV	0,13	0,20



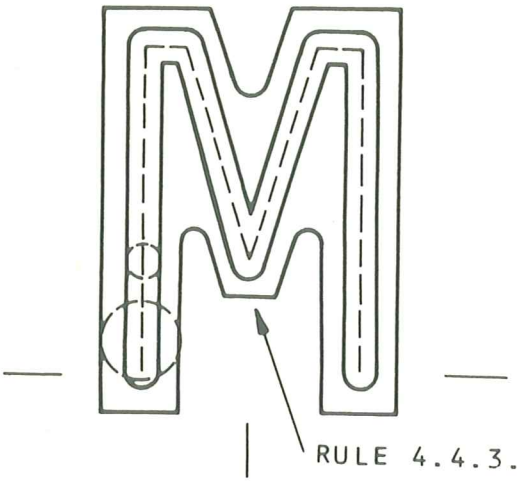
OCR-A
NUMERAL TWO

Fig. 2



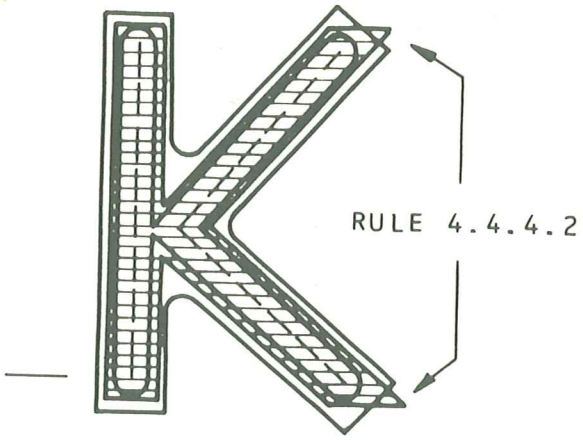
OCR-A
NUMERAL ONE

Fig. 3



OCR-B
CAPITAL LETTER M
CONSTANT STROKEWIDTH FONT

Fig. 4



OCR-B
CAPITAL LETTER K
LETTERPRESS FONT

Fig. 5

4.5 DEFINITION AND MEASUREMENT OF PRINT QUALITY PARAMETERS

4.5.1 GENERAL

For machine recognition, it is essential that the printed image be in accordance with the theoretical shape and that the print contrast signal (PCS) of all parts be sufficiently high, i.e. above a minimum value. This is necessary for the image to be distinguishable from the background. For good reliability of reading it is essential that a majority of the character has a higher PCS value than the minimum which the specification allows for any particular small area portion. Reliability of reading may also be decreased as the unevenness of the printing within the characters is increased.

Whenever possible, the measurements of print quality parameters have been stated both in terms of visual methods as well as instrumented methods. Exceptions are the two parameters "PCS within a Character" and "Contrast Variation within a Character" for which only the instrumented measurement can suitably be described.

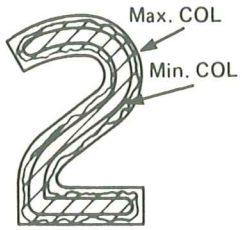
An effort has been made to achieve close agreement between visual and instrumented measurement. Exact correlation is not always possible, and some differences can arise when performing measurements. In case of conflict between the two measurement methods, the instrumented technique should apply.

4.5.2 BEST FIT

All measurements described hereafter are to be made in the "best fit" position of the character with the COL masks.

The best fit can be achieved visually or on an instrument by positioning the actual character image so that it fills the min. COL as much as possible and at the same time does not extend beyond max. COL. More specifically, the overall reflectance within min. COL shall be a minimum. If this condition is met for several positions, then the best fit position is that yielding the maximum reflectance outside max. COL (see A 2.2).

Light portions of the character inside min. COL and dark portions outside max. COL have to be checked as to Edge Irregularities, Cut-Off, Voids and Spots.



Gauge in its "best fit" position

4.5.3 VISUAL MEASUREMENTS

For the visual examination of the character image a transparent gauge is used. The gauge indicates the character outline limits (COL) as defined in section 4.4. The examination has to be done in the "best fit" position.

4.5.4 INSTRUMENTED MEASUREMENTS-BASIC PCS VALUES

Most parameters described hereafter are to be derived from the set of Basic PCS Values obtained as follows:

A gauge for the range required is placed on the character to be measured; this gauge bears the min. COL, the max. COL and the centreline.

An aperture of 0,2 mm diameter is moved along the whole centreline of the gauge in steps of 0,1 mm. All PCS values obtained are recorded in the sequence they have been measured. If the length of the centreline is shorter than 2 mm, the measurements must be made with steps of 0,05 mm (see A 2.3).

4.5.5 PCS WITHIN A CHARACTER

The smallest value of the highest 80 % Basic PCS Values is called PCS_{80} %.

It must satisfy the following conditions:

PCS_{80} % > 0,60 in Range X

PCS_{80} % > 0,50 in Range Y

4.5.6 CONTRAST VARIATION WITHIN A CHARACTER

The variation of contrast within a character is defined by the Contrast Variation Ratio:

$$CVR = \frac{PCS_{max}}{PCS_{min}}$$

where:

PCS_{max} is the highest average PCS value of three consecutive Basic PCS Values for characters with a centreline longer than 2 mm and of five such consecutive values for characters with a centreline of less than 2 mm.

PCS_{min} is the lowest average PCS value of three consecutive Basic PCS Values for characters with a centreline longer than 2 mm and of five such consecutive values for characters with a centreline of less than 2 mm.

CVR must satisfy the following conditions:

CVR < 1,50 in Range X

CVR < 1,75 in Range Y

4.5.7 VOIDS

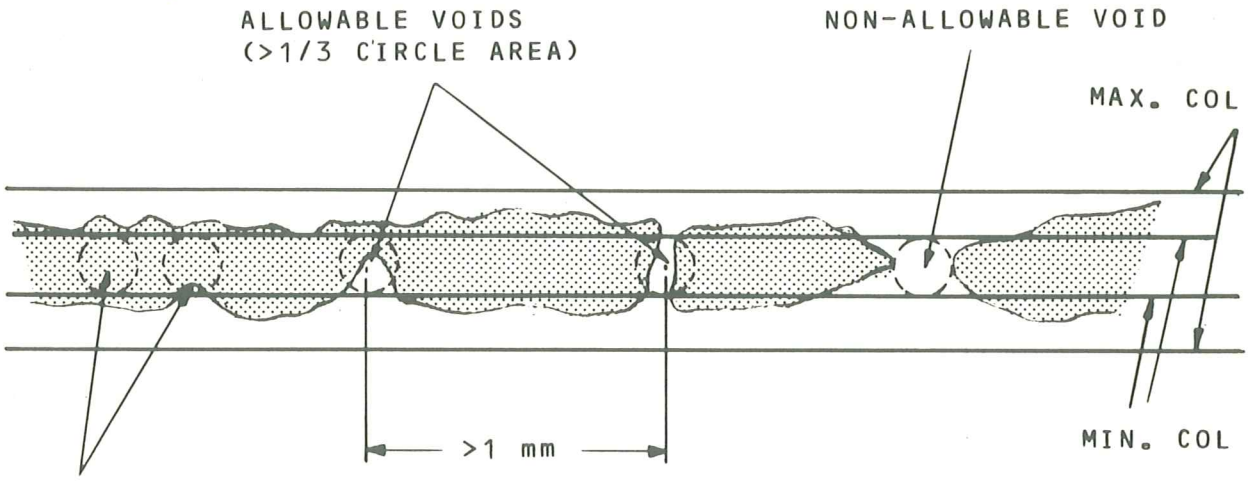
Voids are areas inside the minimum COL which have a significantly lower density than the printed image. Voids can be measured either visually or in terms of PCS. The distinction between allowable voids and non-allowable voids is based on a measurement of their size and distance. Small voids are permitted providing certain conditions are satisfied, larger ones are not (see A 2.6).

4.5.7.1 VISUAL IDENTIFICATION

Any area within the minimum COL is a void if it is printed very lightly in comparison to the rest of the character.

One or more voids are allowable if contained entirely in an inspection circle of 0,2 mm diameter and if their total surface is smaller than one third of the surface of the circle.

If their total surface is greater than one third of that of this circle, then the distance of the centre of this circle to the centre of the inspection circle of the nearest void or group of voids of a total surface greater than one third of that of its inspection circle, must be at least 1,0 mm.



VOIDS ALLOWED IN UNLIMITED NUMBER (<1/3 CIRCLE AREA)

4.5.7.2 INSTRUMENTAL IDENTIFICATION

For the instrumented identification all Basic PCS Values lower than PCS_{80} % are to be considered. The values d mentioned hereafter are:

d = 0,40 in Range X

d = 0,35 in Range Y

A void is present at points for which a PCS of less than d is measured.

a) Characters with a centreline of more than 2 mm

- If a point has a PCS < d and if both adjacent points have a PCS > d, an allowable void is present at this point.

- If two adjacent points have a PCS < d, an allowable void is present only if their distance to the next similar pair of points is at least 11 steps.

- Three or more consecutive points with a PCS < d define a non-allowable void.

b) Characters with a centreline of less than 2 mm

- Single points or pairs of two consecutive points having a PCS < d define allowable voids.

- Groups of 3 or 4 consecutive points having a PCS < d define an allowable void only if their distance to the next similar group of points is at least 21 steps.

- Groups of 5 or more consecutive points with a PCS < d define a non-allowable void.

4.5.8 STROKE EDGES

The nominal strokewidths and their tolerances are given in Table 2. Measurements of the strokewidth and of the character geometry require examination of the edges of the strokes.

Table 2. STROKEWIDTH - NOMINAL VALUES AND TOLERANCES

Size	Heights (H) mm	Nominal Strokewidth mm	Range X Tolerances mm	Range Y Tolerances mm
I	2,40	0,35	± 0,08	± 0,15
III	3,20	0,38	± 0,08	± 0,18
IV	OCR-A: 3,80 OCR-B: 3,60	0,51 0,50	± 0,13	± 0,25

According to ECMA-11 (11.4.1) the nominal value of the strokewidth for the small letters and the characters %, # and @ is 0,31 mm for Size I and 0,44 mm for Size III.

4.5.8.1 VISUAL INSPECTION OF THE STROKE EDGES

The stroke edge is taken at those points where the reflectance is approximately half between that of the nearby area of the stroke and the background.

4.5.8.2 INSTRUMENTED INSPECTION OF THE STROKE EDGES

From the Basic PCS Values, $PCS_{average}$ is derived as the arithmetic means of the highest 80 % Basic PCS Values. The stroke edges are considered within specification if, when moving an aperture of 0,2 mm diameter in steps of 0,2 mm along the min. COL and then along the max. COL, the values obtained along the min. COL are always greater than $0,5 \cdot (PCS_{avg})$ and those obtained along the max. COL are always less than $0,5 \cdot (PCS_{avg})$. See also A 2.5.

If these conditions are not met, the stroke edges exceed one or both COLs. In this case the character must be checked with regard to edge irregularities.

4.5.9 EDGE IRREGULARITIES

An edge irregularity is a point for which the measurements described in 4.5.8.2 produce a value which is either less than $0,5 \cdot (PCS_{avg})$ along the min. COL or greater than $0,5 \cdot (PCS_{avg})$ along the max. COL. An edge irregularity is allowable only if it is at least 1 mm away from another edge irregularity.

4.5.10 SPOTS

Spots are areas outside the maximum COL which are contrasting with the background. They can be measured either visually or in terms of PCS. The distinction between allowable spots and non-allowable spots is based on a measurement of their size. Small spots are permitted providing certain conditions are satisfied, larger ones are not. Spots may be connected or adjacent to parts of the printed image, or may be free standing (see A 2.6).

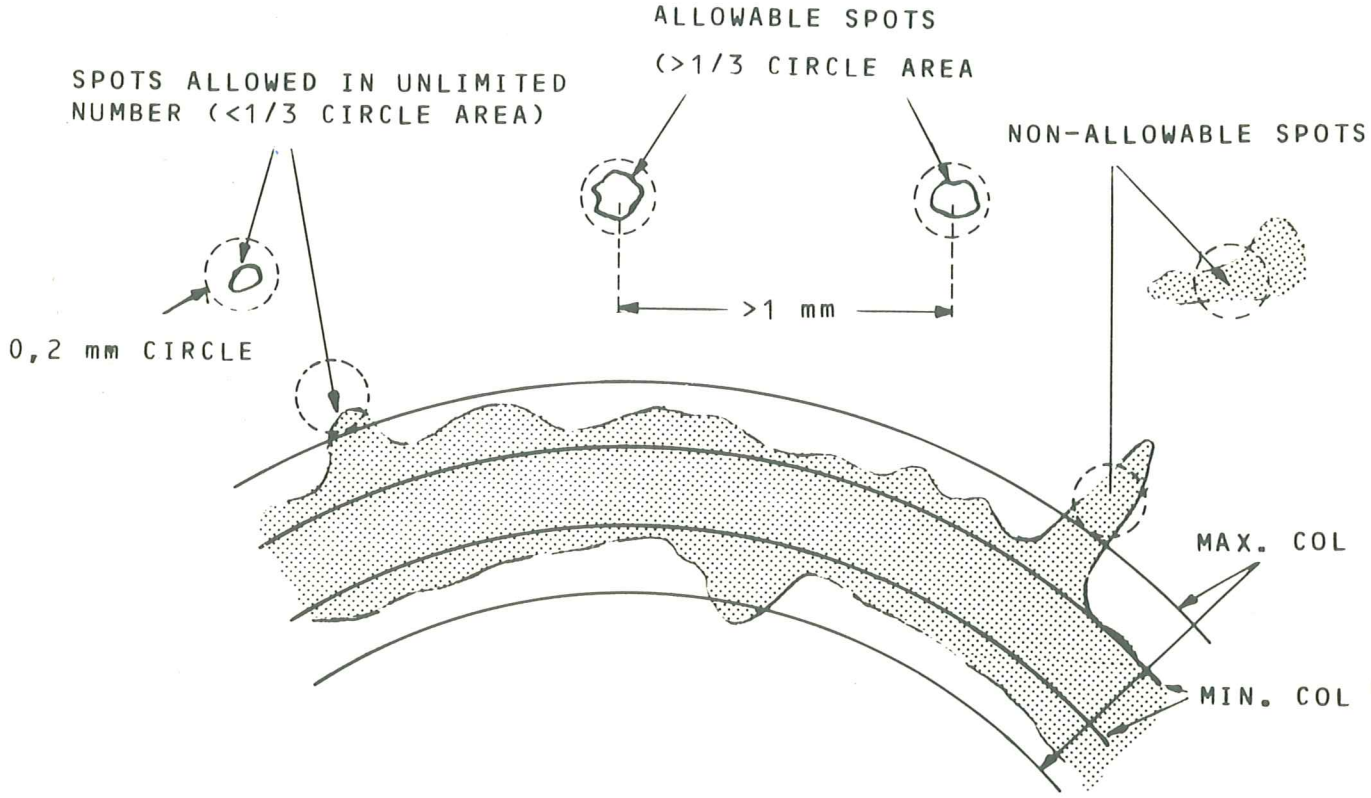
4.5.10.1 VISUAL IDENTIFICATION

When a measuring gauge is best fit to a character, any dark area outside the maximum COL is a spot. The size of a spot determines if it is allowable or non-allowable.

Any spot that is nearly as dark or darker as the lightest printing within the minimum COL has to be checked.

One or more spots are allowable if contained entirely in an inspection circle of 0,2 mm diameter and if their total surface is smaller than one third of the surface of the circle.

If their total surface is greater than one third of that of this circle, then the distance of the centre of this circle to the centre of the inspection circle of the nearest spot or group of spots of a total surface greater than one third of that of its inspection circle, must be at least 1,0 mm.



4.5.10.2 INSTRUMENTAL IDENTIFICATION

Spots are measured with an aperture of 0,2 mm diameter. It is centred on the spot at the point with the highest PCS value. When this position is identified the eight positions defined by the steps of 0,1 mm horizontally and vertically are also measured.

The values e mentioned hereafter are:

e = 0,65 . PCS_{min} in Range X

e = 0,70 . PCS_{min} in Range Y

After measurement of the nine positions mentioned:

- If at least three positions have a PCS > e, the spot is not allowable.

- If at most one position has a PCS > e the spot is allowable.

- If two positions have a PCS > e, the aperture is centred on the position with the smaller PCS and the seven remaining positions defined by steps of 0,1 mm horizontally and vertically are also measured:

  - If a 3rd position is found with a PCS > e, the spot is not allowable.

  - If no 3rd position with a PCS > e is found the spot is allowable only if its distance to a spot of the same type and to the max. COL is greater than 1 mm.

If in this procedure one or more positions happen to have their centre within max. COL, these positions are to be disregarded.

Spots remote from the character, i.e. outside the area of interest, are not subject to PCS limitations. However, if located in the Clear Area (section 5.10), their size is limited to 0,2 mm in diameter.

5. CHARACTER POSITIONING

5.1 GENERAL

Character positioning specifications are needed to ensure that each OCR character is read by the reading device without interference from other OCR characters or from non-OCR matters.

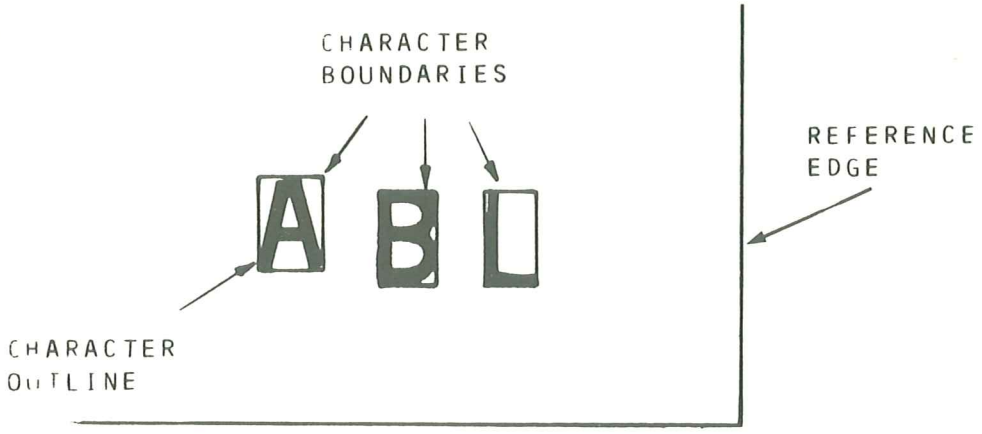
This section contains basic specifications relating to the position of characters on a form to accomodate general requirements of OCR devices. It does not contain all the rules which may be necessary for a particular application. These additional rules will be subject of other standards or publications.

5.2 DOCUMENT REFERENCE EDGES

A number of specifications in this section relate to the document reference edge. These can be horizontal and/or vertical edges.

5.3 CHARACTER BOUNDARY

The character boundary is the smallest rectangle that has one side parallel to the document reference edge and which contains a character when aligned at the mean stroke edge. The mean stroke edge serves as an imaged averaging line across the generally blurred transition area between the printed character and the paper.

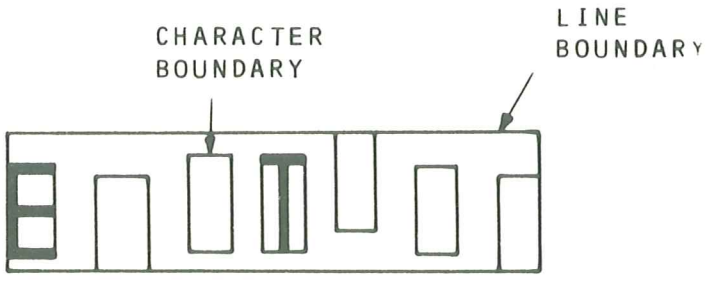


5.4 CHARACTER SKEW

The skew of a character is the rotational deviation of the printed image from the intended orientation relative to a document reference edge. Character skew shall not exceed 3^{o}.

5.5 LINE BOUNDARY

A line boundary is the smallest rectangle that has one side parallel to the document reference edge and which contains all the character boundaries of the characters of the line.



5.6 FIELD

A field is a specific portion of a line which comprises at least one character that may be treated as a unit of information. A line could comprise several fields. Dimensional specifications on field do not appear in this standard.

5.7 HORIZONTAL POSITIONING OF CHARACTERS WITHIN A LINE

5.7.1 CHARACTER SEPARATION WITHIN A LINE

Character separation within a line is the horizontal spacing between the two opposite vertical sides of the character boundaries of two characters within the same line boundary.
The character separation shall not be less than the nominal stroke width for the appropriate size as specified in section 4.

5.7.2 CHARACTER SPACING WITHIN A LINE

Character spacing is the horizontal distance of the vertical centrelines of the character boundaries of two characters within the same line boundary corrected by the distance which would exist between these vertical centrelines, 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. Character spacing shall not be less than :

2,30 mm for character sizes I and III

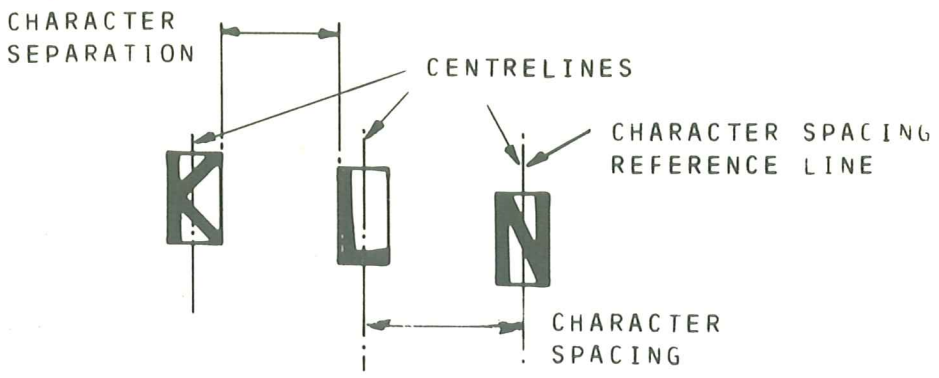
3,30 mm for character size IV

Two characters are adjacent if they belong to the same field and if the character spacing is less than:

4,60 mm for character sizes I and III

6,60 mm for character size VI

Some printing devices like letterpress, variable pitch typewriters and some journal tape printers produce printing that does not meet the character spacing specification for all combinations of characters within the repertoire of the printer. Some OCR scanners can permit this exception as long as the character separation requirements of section 5.7.1 are satisfied. When considering the installation of an OCR system of this type close liaison with printer and reader manufacturer is advised.



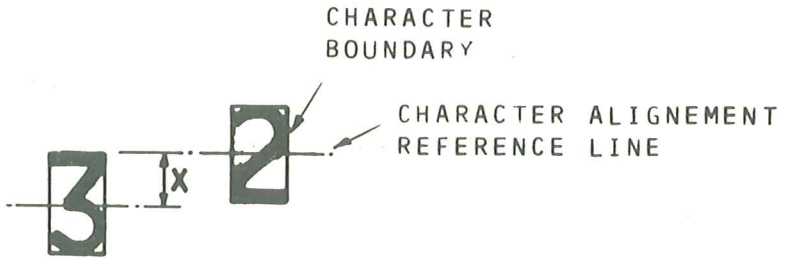
5.8 CHARACTER ALIGNMENT WITHIN A LINE

Character alignment is the vertical distance between the lower side of a character boundary containing one character and the elongated lower side of a character boundary containing another character within the same line boundary corrected by the vertical distance which would exist between the lower side of the character boundaries if the same two characters were superimposed in their nominal position. This definition does not apply to the Long Vertical Mark (see section 5.8.3).

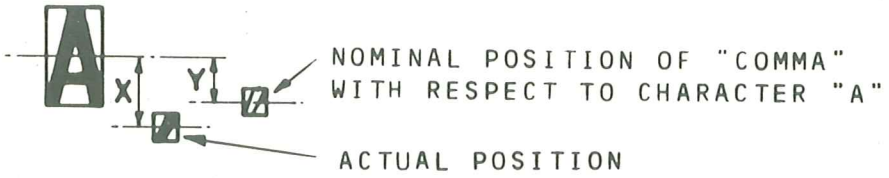
5.8.1 ADJACENT CHARACTER ALIGNMENT

Adjacent character alignment is measured according to the above procedure. It shall not exceed :

0,65 mm for character size I
0,90 mm for character size III
1,10 mm for character size IV



X = ADJACENT CHARACTER MISALIGNMENT



(X-Y) = ADJACENT CHARACTER MISALIGNMENT

5.8.2 CHARACTER ALIGNMENT WITHIN A LINE

Character alignment within a line is measured according to the above procedures. It shall not exceed:

1,30 mm for character size I

1,80 mm for character size III

2,20 mm for character size IV

5.8.3 PRE-PRINTED LONG VERTICAL MARK ALIGNMENT

The PRE-PRINTED LONG VERTICAL MARK should preferably extend beyond the top and the bottom boundaries of any neighbouring character within the same line. This means that if the minimum height of 3,7 mm (see ECMA-11, clause 10) is selected, a good control of misalignment within the line must be achieved. Else it is recommended to select a greater height for the PLVM. This requirement does not need to be met for small letters with descenders.

5.9 PRINTING AREA

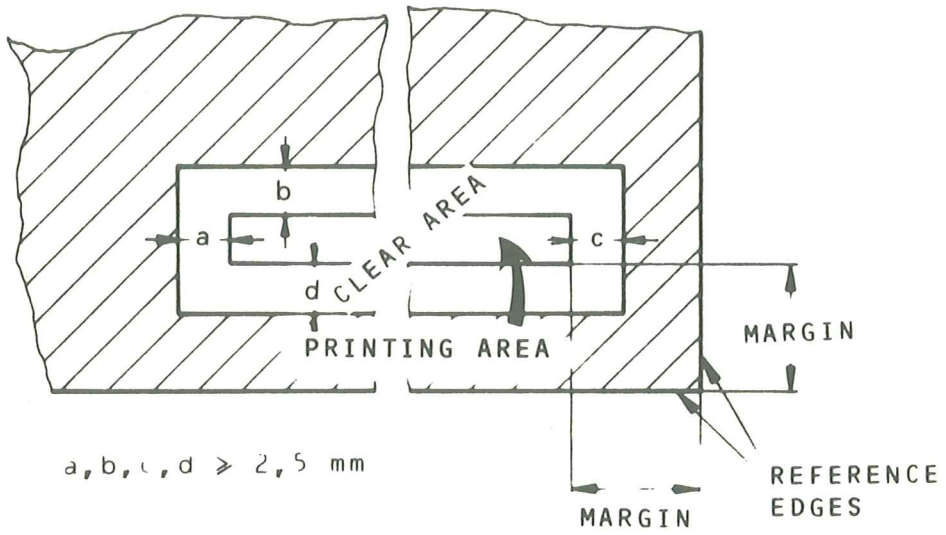
The printing area is a rectangle that has one side parallel to the document reference edge and is intended to contain only machine readable characters of one line. The line boundary of a line of printed characters must completely be inside the printing area.

5.10 CLEAR AREA

A clear area is defined as that region of a document reserved for one line of OCR characters and the clear space around these characters. The clear area is placed symmetrically around the printing area. The locations and dimensions of clear areas will be determined by the nature of the individual applications and the requirements specified in this section. The distances a, b, c and d between the corresponding boundaries of the printing area and the clear area should not be less than 2,5 mm. For readers able to read several lines on the same document a number of clear areas and print areas is defined on the document. For this type of readers section 5.12 and 5.13 apply. For two succeeding lines the clear areas of the two lines may overlap (or the clear space between the lines may be shared).

5.11 MARGIN

The distance between any boundary of the printing area and the nearest parallel paper edge is called the margin. Normally a margin shall be at least 6,35 mm. Where manually operated serial entry devices (e.g. typewriters) are used, it is recommended to use top and bottom margins of at least 25,4 mm.



5.12 LINE SEPARATION

Line separation is the vertical distance between the upper line boundary for a line of print, and the lower line boundary for the line immediately above.
The parameters which influence line separation are line pitch specification, line skew, vertical alignment, character height and stroke width.
In any case the minimum line separation shall not be less than:

0,65 mm for character size I
1,50 mm for character size III
2,00 mm for character size IV

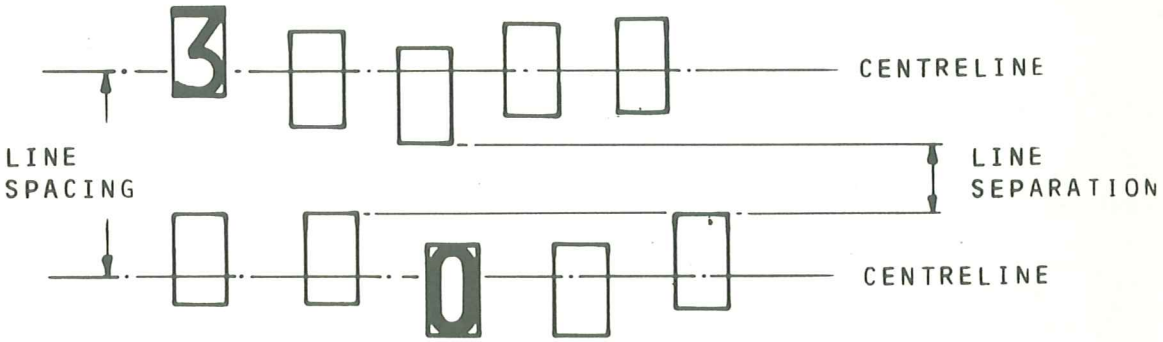
If the character sizes are intermixed the line separation limitation for any pair of lines shall be that applicable to the largest character in the two lines.

5.13 LINE SPACING

Line spacing is the vertical distance between the average horizontal centreline position of all characters printed on one line and that of all OCR characters printed on the next line.
In any case, the line spacing shall not be less than:

4,20 mm for character size I
4,80 mm for character size III
5,30 mm for character size IV

If character sizes are intermixed the limitation applying to the largest size applies. When lower case size I characters are being used, line spacing shall not be less than 4,80 mm.



ANNEX

A 1 PAPER CHARACTERISTICS AND MEASUREMENTS

A 1.1 Spectral Properties

A 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 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.

A 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 paper, 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.

A 1.1.3 Notes on Measurements

A 1.1.3.1 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.

A 1.1.3.2 Fluorescent Additives

While a low level of fluorescence may be unavoidable due to recycling of paper, efforts should be made to minimize this contamination, and fluorescent additives should generally be excluded in the manufacture of paper made for OCR use.

This is necessary both to avoid difficulties in reading (with particular equipment) and in sorting (where fluorescent materials are added by the user). It is also recognized that other readers can tolerate fluorescent additives deliberately included for identification purposes.

A 1.2 Paper Opacity

A 1.2.1 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 transport system that provides a non-uniform backing surface. For such cases a more opaque paper should be used, or higher PCS value should be required for OCR information.

A 1.2.2 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 have 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 70 % 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.

A 1.2.3 Variation of Paper Reflectance

A 1.2.3.1 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 in the order of 0,1 mm in diameter constitutes the input to a photo-detector. 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 fiber 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 paper and the printing. Embossment of patterns or the printing of coloured patterns should be strongly avoided.

A 1.2.3.2 Measurement

A 1.2.3.2.1 Reflectance of the Black Backing

The backing of 0,5% reflectance is most easily provided by an unlit cavity. If it 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.

A 1.2.3.2.2 Procedures

It appears that 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 they can be considered as being non-correlated. The number of observations required for a 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 be covered, for high or medium opacity papers respectively, in order to avoid the influence or correlations 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:

- Arrange the measurements obtained by one of the above scanning methods in a descending order of magnitude.
- Exclude highest 0,5% and the lowest 0,5% of the values. Calculate the ratio Rmax/Rmin for the remaining values.
- This ratio should not exceed 1,2 for the high opacity class and 1,3 for the 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.

A 1.3 Paper Gloss

A 1.3.1 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 that 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.

A 1.3.2 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.

A 1.4 Dirt in Paper - Grid Assay Method

A 1.4.1 Equipment

This should consist of the following:

Grid	- A frame 1 m x 1 m divided into 100 squares by fine wire.
Working surface	- To take paper and frame for viewing at a distance of about 0,5 m.
Lighting	- Even, high intensity northlight illumination of the working surface.

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

Timer - To indicate ½ or 1 min. intervals.

Counter - To tally the number of squares containing dirt.

A 1.4.2 Sampling and Test Area

A total of 6 m^2 to represent the reel or stack of sheets. Reels to be sampled at both ends with 6 x 1 m samples representing the full width for mills reels, (sampling from the outer end of the preceding reel in manufacturing sequence if necessary). Sheet stacks to be sampled in 6 positions with sufficient sheets to make up the total area.

A 1.4.3 Method

Lay out sample topside uppermost.

Remove loose dirt and dust from surface.

Place grid over sample.

Start timer and scan all the squares in turn in 1 minute. Recording 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 and record as the number of squares containing dirt/6 m^2.

A 1.4.4 Note on Calibration

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.

A 1.5 Lighting Standard

A 1.5.1 Preface

The following conditions are taken from BS 950: Part 2: 1967 Artificial daylight for the assessment of colour - Part 2
Viewing conditions for the graphic arts industry. Currently there is no ISO equivalent. The spectral distribution of the illumination and the viewing conditions are specified.

A 1.5.2 Foreword

Illuminants with correlated colour temperatures from 7400°K to 3000°K have been suggested as standards. In this standard, a form of daylight at a correlated colour temperature of 5000°K has been selected, which conforms to the latest CIE Recommendations, and has a nominal chromaticity of

$$x = 0,3457 \quad y = 0,3586$$

This correlated colour temperature has been chosen in preference to higher or lower values to give good discrimination at the red and blue ends of the spectrum and to minimize difficulties due to metameric effects.

The level of illumination, and the geometry of lighting and viewing are specified, but no specification can be laid down for the light source itself, because

a) the light from the source may be modified by the apparatus, and

b) the required quality of lighting may be obtained from a mixture of the light from different sources.

Conditions to be ovserved with respect to surround lighting are also laid down.

A 1.5.3 Illuminant

The light falling on the viewing surface shall be of the composition specified in Table 1, subject to a tolerance of ± 15 %; where two contiguous bands deviate in the same sense the mean deviation of the two bands shall not exceed 7,5 %.

TABLE 1 SPECTRAL DISTRIBUTION (VISIBLE LIGHT)

Spectral band	Wavelength range	Band value for 100 lumen flux (5000° K)
	nm	lm
1	400 - 455	0,573
2	455 - 510	9,6
3	510 - 540	21,8
4	540 - 590	44,2
5	590 - 620	15,8
6	620 - 760	8,01

Fluorescent brighteners as commonly occuring in white papers have negligible effect on colour reproduction.

A 1.5.4 Levels of Illumination and Luminance

The average level of illumination for viewing shall be 2200 ± 470 lx and of such degree of uniformity that, if a uniform sheet of white paper is placed in the viewing position, it shall appear uniformly illuminated.

Disturbing glare from light sources and reflecting surfaces should be avoided.

A 1.5.5 Surround

Any external lighting, including that from reflecting surfaces, shall not interfere with the nature and intensity of the viewing light so as to cause the limits specified for this by A 1.4.3 and A 1.4.4 to be exceeded, and it shall not interfere with the state of visual adaptation of the observer. Any uncovered areas of viewing surfaces shall be mid-grey.

A 1.5.6 Geometry of Lighting and Viewing

The light shall be so incident and the reflecting surface shall be viewed in such a manner that the effects of specular reflection are minimized.

A 1.5.7 Design of Practical Installations

As specialised lighting knowledge is required, this section is provided to assist in the selection of equipment.

Lighting meeting the requirements of A 1.4.3 may (with the exeption of the ultra-violet bands) be achieved in practice by mixing the light from tungsten filament lamps operating at a colour temperature of 2400°K with that from fluorescent tubes of the light blue type which have peak visible output at about 475 nm. The mixture must be in equal luminous proportions.

As a guide, 240 V, 60 W tungsten filament vacuum rough service lamps have a colour temperature of approximately 2400°K and a luminous efficiency of 8 lm/W; light blue fluorescent tubes have an efficiency of approximately 40 lm/W average through life. Hence assuming equal utilization of light output the ratio of power required by tungsten lamps to those required by fluorescent lamps is 5 : 1.

Precise figures cannot be given as these will depend on the design of the fittings.

It is important that each lighting fitting produces a correct mixture of light from the fluorescent tubes and tungsten filament lamps, and that all transmitting and reflecting components are neutral in colour. The two types of light source should be so arranged as to maintain light of unchanging spectral composition over the viewing plane. The light output of fluorescent tubes varies with the glass wall temperature. Consequently adequate ventilation should be provided and radiant heat from adjacent filament lamps must be minimized. Metallic sheeting, painted with a matt and neutral finish, or anodized aluminium sheeting may be used to lessen heat radiation between lamps whilst still maintaining a high light output from the fittings.

To adjust for evenness of illumination and correct spectral composition the light sources should be switched on independently and the level of illumination provided by the light blue fluorescent tubes made equal to that from the tungsten filament lamps. In making the measurement it should be noted that illumination meters usually require correction to match their spectral response to that of the normal human eye.

Specular reflection of the light source in the direction of the observer should be minimized in order to reduce the effects of

(1) imaging of the light source,

(2) discomfort glare,

(3) desaturation of colours.

Lighting from above involves elaborate optical design of the lighting fittings if the viewing surface is to be evenly illuminated and viewed normally. Tilting the viewing surface only serves to increase the effects of specular reflection.

Most of the difficulties of the overhead arrangement can be overcome by placing the lighting fittings vertically at either side of an almost vertical viewing surface.

Apparatus to meet the requirements of this British Standard requires careful design. One way is to use simple reflectors pivotable about their vertical axes and suspended from horizontal radius arms attached to a wall behind the viewing surface. To avoid the introduction of objectionable specular reflection, a stop should be provided to prevent the fittings from being swung too far forward.

SPECTRAL POWER DISTRIBUTION OF ARTIFICIAL DAYLIGHT (5000°K)

Values derived from 1965 CIE Recommendations			
Wavelength	Relative power per unit Wavelength	Wavelength	Relative power per unit Wavelength
nm		nm	
300	0,02	550	102,3
310	2,0	560	100,0
320	7,8	570	97,7
330	14,8	580	98,9
340	17,9	590	93,5
350	21,0	600	97,7
360	23,9	610	99,3
370	26,9	620	99,0
380	24,5	630	95,7
390	29,8	640	98,8
400	49,3	650	95,7
410	56,5	660	98,2
420	60,0	670	103,0
430	57,8	680	99,1
440	74,8	690	87,4
450	87,2	700	91,6
460	90,6	710	92,9
470	91,4	720	76,8
480	95,2	730	86,6
490	92,0	740	92,6
500	95,7	750	78,2
510	96,6	760	57,7
520	97,1	770	82,9
530	102,1	780	78,3
540	100,8		

A 2 CHARACTERISTICS OF THE PRINTED IMAGE

A 2.1 General

This standard 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 assure maximum compliance with the standard. 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.

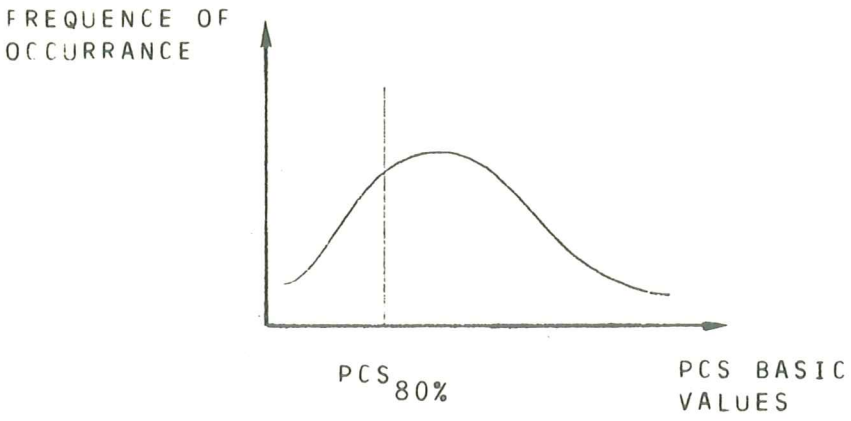
A 2.2 Best Fit

The definition of the best fit allows for its determination with high accuracy by the instrumental method. With the visual method, different operators will not select identical positions. Tests have been conducted, during which operators of different companies measured the same samples. These tests have shown that slight differences in the selection of the best fit position lead only to negligible differences between the values of the parameters derived from the set of basic values measured in the different best fit positions and those obtained for the same samples by the instrumental method. With other words, it appears that the selection of the best fit position for the visual method is not a critical operation with regard to the reproducibility of the measurements.

A 2.3 Basic Values

Most print quality parameters are derived from the PCS basic values measured as specified in 4.5.4. These basic values depend for each printed character on the starting point selected along its centreline. The tests mentioned above have shown that the print quality parameters are not affected by the choice of this starting point.

This is due to the fact that all print quality parameters are obtained as the average of at least 3 basic PCS values (see for example PCS_{min} and PCS_{max}). PCS 80 % too is not an isolated basic PCS value, but it is a limit value between the highest 80 % and the lowest 20 % of points (in statistical language it is a "quantile") as shown in the figure below.



A 2.4 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 visible or the near IR spectrum.

A printing ink provides good absorbance in one or both 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 characteristic 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 both 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.

A 2.5 $PCS_{Average}$

For a rapid inspection of stroke edges PCS_{AV} can be approximated by :

$$\frac{PCS_{max} + PCS_{80\ \%}}{2}$$

For a rigorous assessment PCS_{AV} must be calculated as indicated in 4.5.8.2.

A 2.6 Spots and Voids

A 2.6.1 Definition

A printed image contains, in most cases, voids within the minimum COL and spots outside the maximum COL but within the neighbourhood of the characters. These spots are defined as character associated spots.

A 2.6.2 Significance of Spots and Voids

For machine recognition of the printed image it is essential that the print intensity of all parts should be high enough to exceed a certain minimum value and be distinguishable from the background. These requirements are covered by the specifications for voids and spots.

A 2.6.3 Visual identification

The visual methods defined in 4.5.7.1 and 4.5.10.1 for the identification of spots and voids rely on the observer's estimation of the area and the reflectance of the void or spot. Whilst estimate of the area may easily be made, it is more difficult to assess accurately the contrast of the spots and voids. Therefore great care must be taken in making these visual examinations.

A 2.6.4 Instrumental Identification

The minimum PCS found within the outline of a character is a measure of the smallest useful signal that the character will produce in an OCR scanner. If the detection threshold is put above this value, the character will display voids.

Because of the distinction between allowable (small) voids and non-allowable ones, the specification for voids is, in general, somewhat higher than this minimum. It is defined by the PCS threshold d above which all voids are considered allowable. Broadly speaking, it is a measure of the contrast between the character and its background. The specification for spots likewise, is not the PCS level at which spots first appear but the threshold level e beyond which they are considered too large to be allowable. It is related to the intensity of background noise in the region of the character.

The values d and e have been defined to take account of the different requirements for voids and spots for the print quality ranges X and Y.

As the incidence of voids increases, the print contrast diminishes until in the limit it is no greater than the level of reflectance irregularities in the paper. A decrease in voids will tend to improve reading system performance. This can be achieved, for instance, at some extra cost, by a reduction in the allowed duration of ribbon life.

A 2.7 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.

A 2.8 Spots remote from a character

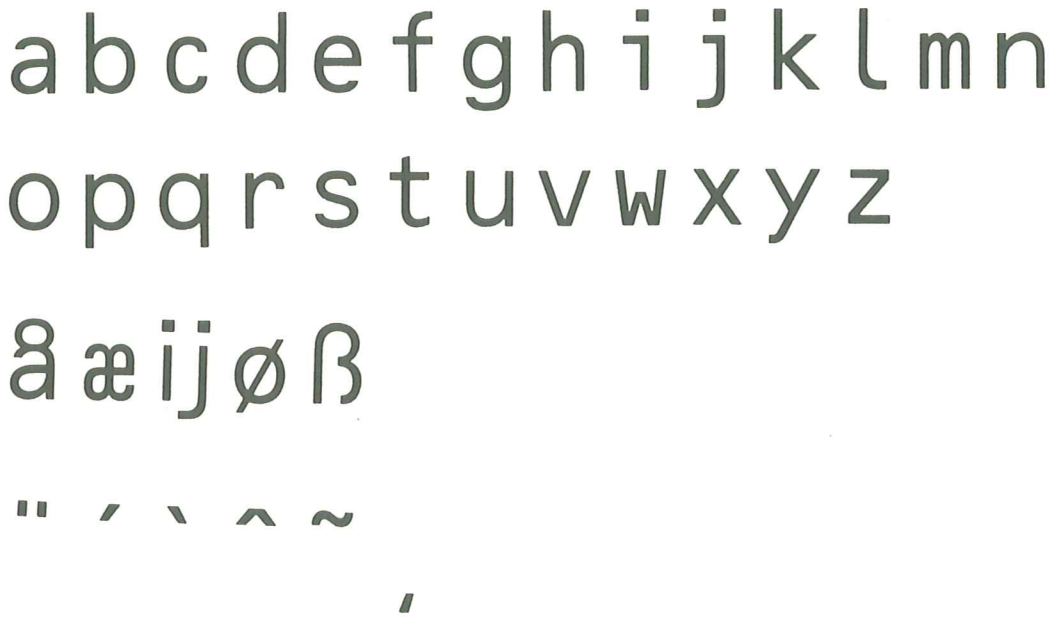
The area of interest of a character is defined in 4.3.1 as an area twice the nominal character height by twice the nominal character width and centred on the character being measured. The PCS level and frequency of spots in the area of interest of a character are specified in 4.5.10.2.

The size and frequency of spots remote from a character should also be strictly controlled. The size of spots should be minimized; printing smudges and regular patterns of dots should be avoided.

Many reading operations are started upon detection of the first black point and if a spot occurs larger than 0,2 mm, then the recognition process may begin. It is advisable that spots greater than 0,2 mm be prevented.

A 2.9 Recommendation for lower-case OCR-B characters

For the following set of characters a higher print quality is required, both in terms of PCS and strokewidth. Strokewidth variations should be maintained within range X.



A 3 CHARACTER POSITIONING

A 3.1 Objectives of the character positioning requirements

Character positioning specifications (Format Rules) are needed to ensure that each OCR character on a document is "seen" by the reading device without interference from the other OCR characters or from non-OCR matter. The Format rules given in the standard (which are explained in the following sections) are to be taken as minimum requirements and may need to be supplemented by further rules for specific systems.

A 3.2 Document Reference Edge

The document used in an OCR system must be moved and suitably positioned for printing and reading the OCR information. One or more document edges are used to provide a reference for these operations. Because of the diverse nature of OCR documents, it may sometimes be convenient to specify one reference edge (e.g. for journal rolls); for others it may be necessary to specify two edges (e.g. for cheques the bottom right hand edges are usually specified).

The tolerance on the distance between the average horizontal centreline on a line of OCR characters and a top or bottom reference edge may be vital to the satisfactory functioning of the system. No dimension for this tolerance is given in the specification, since system requirements differ widely, but its importance must not be overlooked.

A 3.3 Clear Area, Printing Area and Margin

OCR printing must be isolated from all other printing or patterns on the document in order to allow the reading device to distinguish the OCR information more readily. This isolation is provided by maintaining a "border" of blank paper between the OCR information and the remainder of the document. From this arises the distinction between the "Printing Area", which must include all of the OCR characters and the larger "Clear Area" which includes the Printing Area and must be free from any other printing or embossing. If the distance between the boundary of the Clear Area and that of the Printing Area approaches the minimum specified, due account must be taken of printing tolerances (vertical misalignment, etc.) and expected paper dimensional changes. It is good practice in document design to provide as generous a Clear Area as possible.

The boundary of the Printing Area should be kept well within the paper edges, i.e. the margins should be large. This has the advantage, among others, that a moderate degree of edge mutilation can take place without impairing readability. There are some special cases however, where the small size of the document may make large margins impracticable and the boundary of the Printing Area may then have to lie close to the document edge(s). Relaxation of the specification in this respect is only permissible when it has been established that all OCR devices in the system can handle these documents.

The dimensions of the printing area and its position relative to the document edge(s) are important for readers that have limited line finding capabilities.

A 3.4 Line Spacing

Line spacing is only significant for multiple line documents. It is the intent of the standard to limit the number of lines of printing that may occur within a given vertical distance. This limitation is necessary in addition to the requirements of line separation, since characters in a line may all be less than full character height, (e.g. symbols, such as minus). In such cases, the line spacing must be maintained to permit printing of full height character.

The maximum line packing density permitted by the tolerances given in the body of the standard for the four character sizes are approximately:

Size	I	III	IV
Lines per 25,4 mm	6	4	3

However, for these values to be acceptable the tolerances on the parameters influencing Line Separation must be below the maxima specified, which apply for wider spacing. (The parameters which influence Line Separation are: line pitch tolerance, vertical misalignment, character height and strokewidth).

In general, line spacing should be kept as large as possible consistent with the other requirements of the system.

A 3.5 Line Separation

Line separation defines the isolation required between successive lines of OCR information.

Some documents may require and permit more dense spacing of lines of OCR information that can be accomodated with the recommended line separation of 2,5 mm. See A 3.4 above.

An absolute minimum value of line separation for each of the four character sizes is given. Where this minimum is approached an effort should be made to ensure as large a Line Separation as possible, by controlling character alignment, character strokewidth, and if possible, line spacing.

A 3.6 Character Boundary and Character Reference Lines

The Character Boundary is defined for the actual printed image under examination rather than for an ideal character. This is done in order that the limits assigned to the separation between characters and lines shall be realistic and applicable to any quality of print.

In addition to the Character Boundary, which varies in extent according to the strokewidths (wide or narrow), it is necessary to specify vertical and horizontal reference lines, which depend only on the position of the printed character and which serve for measurements of character spacing and misalignment.

Two alternative definitions of the character alignment reference line are given. In marginal cases where the two methods may yield different results, the one giving the smaller misalignment is to be accepted.

A 3.7 Character Spacing

It is the object of the character spacing requirement of the standard to define the lateral relationship of any pair of characters side by side in the same field in such a way that the maximum and minimum character separation requirements can be met.

As mentioned in 5.7.2, the specification on Character Spacing will not be met when variable pitch or variable set width printing is used (e.g. variable pitch typewriters, letterpress). Since these types of printing use wide variation in the character width and spacing they may impose difficulties for OCR devices, special consideration must be given to the compatibility of the print and the reading equipment.

A 3.8 Character Separation

It is a primary requirement of OCR that characters side by side in the same line shall be isolated by a clearance of unprinted paper. This separation constitutes a vertical band (of width not less than the nominal strokewidth, as defined in section 4.5.8) which may not be intruded upon by any part of the character outline.

In order to satisfy the minimum character separation requirement, in difficult cases where the nominal character spacing is close to the minimum, the following points need particular attention:

a) Strokewidth variation

b) Character skew

c) The difference that exists for certain characters between the character spacing reference line and the vertical reference line given in the character drawings. For the Class B character L (size I), for instance, this distance is as much as 0,18 mm .

A 3.9 Character Misalignment

The vertical misalignment of characters should be limited to reduce the cost and complexity of OCR devices, to an extent that is compatible with normal and relatively unsophisticated printing equipment.

The misalignment may be due to:

a) misalignment of individual print faces,

b) misalignment of the document in the printer, causing a complete group of characters printed at one time to be displaced vertically and/or tilted (skew),

c) local distortion or folding of the document before, during or after printing.

This section of the specification limits the degree of misalignment of adjacent characters, with an overall limit on the misalignment of any two characters in a field. The allowable misalignment within a complete line (more than one field) of OCR characters is not specified but will be determined by the reading device(s) in the system.

Misalignment in this case could be the effect of the fields being printed at different times by different printing devices. Therefore it is important to determine the potential misalignment and the requirements for a specific application in order that specifications and controls can be established.

The measurement of actual character misalignment is complicated if symbols, lower case characters, and others which are not of the normal height are present. For such characters the alignment reference lines as defined will be at different levels even when there is no misalignment. Appropriate correction factors can be obtained from the character drawings.

A 4 CHARACTER REPERTOIRES AND PRINT QUALITY RANGES

In correlation with the character repertoire, the following ranges shall be used. Of course, characters of the repertoires stated for range Y may also be printed in range X.

This Standard does not cover characters printed in lower quality range.

OCR-A	OCR-B	Size	Range
	Subset 3 and 4, ECMA-11 (119 char.)	I	X
	Subset 3 and 4, ECMA-11 (119 char.)	IV	
	Subset 2, ECMA-11 (47 char.)	I	Y or X
	Subset 1, ECMA-11 (22 char.)	III	
	Subset 2, ECMA-11 (47 char.)	IV	
Full Set ECMA-8 (14 char.) " " " " " "	Basic Set ECMA-30 (14 char.) " " " " " "	I III IV	Y or X