BS EN
14603:2004

Information technology — Alphanumeric glyph image set for optical character recognition OCR-B — Shapes and dimensions of the printed image

The European Standard EN 14603:2004 has the status of a British Standard

ICS 35.040; 37.080

BSi

British Standards

National foreword

This British Standard is the official English language version of
EN 14603:2004. It supersedes BS 5464-2:1977 which is withdrawn.

The UK participation in its preparation was entrusted to Technical Committee
ISTIECCT/-, Information systems technology international and european
coordination committee, which has the responsibility to:

— aid enquirers to understand the text;

— present to the responsible international/European committee any
enquiries on the interpretation, or proposals for change, and keep the
UK interests informed;

— monitor related international and European developments and
promulgate them in the UK.

A list of organizations represented on this committee can be obtained on
request to its secretary.

Cross-references

The British Standards which implement international or European
publications referred to in this document may be found in the *BSI Catalogue*
under the section entitled "International Standards Correspondence Index",
or by using the "Search" facility of the *BSI Electronic Catalogue* or of
British Standards Online.

This publication does not purport to include all the necessary provisions of a
contract. Users are responsible for its correct application.

**Compliance with a British Standard does not of itself confer immunity
from legal obligations.**

Summary of pages

This document comprises a front cover, an inside front cover, the EN title page,
pages 2 to 33 and a back cover.

The BSI copyright notice displayed in this document indicates when the
document was last issued.

Amendments issued since publication

Amd. No.	Date	Comments

EUROPEAN STANDARD

NORME EUROPÉENNE

EUROPÄISCHE NORM

EN 14603

December 2004

English version

Information technology - Alphanumeric glyph image set for optical character recognition OCR-B - Shapes and dimensions of the printed image

Technologies de l'information - Jeu d'images de glyphe
alphanumérique pour la reconnaissance optique de
caractères OCR-B - Formes et dimensions de l'image
imprimée

This European Standard was approved by CEN on 17 June 2004.

CEN members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration. Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the Central Secretariat or to any CEN member.

This European Standard exists in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CEN member into its own language and notified to the Central Secretariat has the same status as the official versions.

CEN members are the national standards bodies of Austria, Belgium, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland and United Kingdom.

EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPÄISCHES KOMITEE FÜR NORMUNG

Management Centre: rue de Stassart, 36   B-1050 Brussels

Ref. No. EN 14603:2004: E

Foreword

This document (EN 14603:2004) has been prepared by Technical Committee CEN/TC 304, "Information and communication technologies – European localization requirements", the secretariat of which is held by SIS.

This European Standard shall be given the status of a national standard, either by publication of an identical text or by endorsement, at the latest by June 2005, and conflicting national standards shall be withdrawn at the latest by June 2005.

The document is based on the International Standard ISO 1073/II, "*Alphanumeric character set for optical recognition – Part II: Character set OCR-B – Shapes and dimensions of the printed image*".

According to the CEN/CENELEC Internal Regulations, the national standards organizations of the following countries are bound to implement this European Standard: Austria, Belgium, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland and United Kingdom.

Contents

Introduction

Optical Character Recognition technology, OCR, came into use in the 1960s, and some specialized OCR fonts were designed at the time. In 1976 two such fonts were formally standardized by ISO, designated OCR-A and OCR-B, in the standard ISO 1073 parts I and II, respectively.

ISO 1073 was developed by the ISO Technical Committee ISO/TC97, *Computers and information processing.* At the creation of ISO/IEC JTC 1, responsibility for ISO 1073 was transferred to JTC 1/SC 2, *Coded character sets.*

In order to enlarge the set of characters covered by the standard, especially with special letters used in European-origin languages, a revision of the standard was initiated in 1994 by JTC 1/SC 2, and progressed through three consecutive Committee Drafts. Since however testing of the proposed character set extensions could not be accomplished, the JTC 1/SC 2 revision was discontinued in 1999.

With the introduction of the Euro sign a need – primarily European – to add that character to the OCR-B set was recognized. CEN/TC304 therefore decided to develop an OCR-B glyph image shape for the character, verify its recognition properties, and include it in a European version of the OCR-B standard; see CEN/TC304 reports referenced in the Bibliography. The decided-on glyph image shape is specified in Annex A.

For reasons of continuity, and also to facilitate possible future CEN – ISO/IEC cooperation on OCR-B, it was decided to use the current ISO text with only the necessary minimum of changes as a basis for the CEN standard, even though the ISO text was developed in an OCR-technology situation rather different from the one existing when this CEN standard is published. In particular, the ISO standard text's division into clauses was kept as far as possible, although some restructuring might have been desirable.

A description of the main differences between this European Standard and ISO 1073/II is given in Annex B. General information on the implementation of the OCR-B shapes, taken from ISO 1073/II, has been included in Annex C.

In connection with the verification of the recognition properties of the Euro sign, some limited verification was also done on special letters identified during the JTC 1/SC 2 revision work as needed in OCR-B. The extent of this verification is not sufficient for the inclusion of the letters in the OCR-B repertoire at present, but the issue is described in Annex D, as a basis for possible future inclusion work.

1   Scope

This European Standard defines a set of glyph images designated OCR-B, intended primarily for use in Optical Character Recognition (OCR) applications, but suitable also for visual, i.e. human, reading. It does not relate any coding scheme with these images (see clause 5).

This European Standard is based on the ISO standard 1073 part II. It differs from that standard in extending normatively the set of glyph images with the Euro currency sign; but also in deleting some glyphs not relevant in present-day OCR processing. It further adds information on a number of glyph images corresponding to characters specific to some European-origin languages.

NOTE   In ISO 1073 Part II the term "character" is used not only in its strict sense, but also to mean the printed images used for their visual, i.e. printed, representations. In this European Standard the term "glyph image" is used in the latter sense.

This European Standard contains information on nominal dimensions for the glyph images. Tolerances, printing quality and other characteristics of the formats needed to satisfy interchange requirements are covered in other standards (see clause 3).

The glyph image set contains 117 glyph images comprising digits, capital and small letters, diacritical marks, and symbols. It also contains a definition for SPACE.

The diacritical marks are designed for combination with small letters to produce composite glyph images complementing the basic image repertoire.

2   Conformance

A printing or OCR reading device is in conformance with this standard if it can generate/recognize, for either or both of the defined styles (see clause 6) and in one or more of the specified sizes (see clause 7), all or part of the specified glyph image subsets (see clause 9).

A claim of conformance shall specify all the images in (each of) the style(s) and size(s) generated/ recognized. Such a specification shall take the form of a reference to one of the subsets, a list of the images generated/recognized, or a combination of those.

Additionally, a printing or OCR reading device must claim conformance to International Standard ISO 1831 (see clause 3).

Printed images produced by an OCR-B printing device are in conformance with this standard if their nominal shapes and dimensions are in accordance with their respective reference drawing(s) and, in the case of the Euro sign glyph image, with Annex A (see clause 14); with the claimed conformance to tolerances and printing quality factors specified in standard ISO 1831 considered.

3   Normative references

This European Standard incorporates by dated or undated reference, provisions from other publications. These normative references are cited at the appropriate places in the text and the publications are listed hereafter. For dated references, subsequent amendments to or revisions of any of these publications apply to this European Standard only when incorporated in it by amendment or revision. For undated references the latest edition of the publication referred to applies.

ISO 1831-1980, *Printing specifications for optical character recognition.*

ISO/IEC 9541-3:1994, *Information technology — Font information interchange — Part 3: Glyph shape representation*

OCR-B character reference drawings and glyph definition (see clause 14).

4   Terms and definitions

For the purposes of this European Standard, the following terms and definitions apply:

4.1
character
a member of a set of elements used for the organisation, control or representation of data.

4.2
coded character set
a set of characters, defined by unambiguous rules that establish the character set and the relationship between the characters of the set and their coded representations.

4.3
composite glyph image
An image printed on paper or any other medium intended for OCR applications, obtained by superimposing two or more glyph images on the same area.

4.4
glyph
A recognizable abstract graphic symbol which is independent of any specific design.

4.5
glyph image
An image of a glyph, as obtained from a glyph representation printed on paper or any other medium intended for OCR applications.

NOTE   The definition above of "coded character set" differs slightly from definitions in ISO/IEC standards, and the definition of "glyph image" is more limited. The definition of "composite glyph image" is specific to this standard (at the time of its publication).

5   Coding in OCR applications

This standard defines a set of glyph images, but does not specify corresponding characters, and relates no coding with the images. The images have been named as far as possible in the same way as the characters with corresponding glyphs in the ISO/IEC standard 10646-1 (see Bibliography), but this does not imply any normative association between the OCR-B glyph images according to this European Standard and the characters of either ISO/IEC 10646-1 or any other standard for coded character sets.

Printing and/or OCR applications based on this European Standard must therefore define, through reference to other standards or otherwise, the set of glyph images which is available for printing and/or shall be recognized, and for each image the corresponding character and its coding.

6   OCR-B styles

The OCR-B glyph images are defined by this standard in two different styles.

The "constant-strokewidth" style is intended primarily for printer equipment in which the width of the strokes of the images is less controllable. This is for instance the case for some types of mechanical printers.

The "letterpress" style is intended for printing equipment which can reproduce fine details with high accuracy. For aesthetic reasons, the strokewidths of the letterpress images are varied deliberately, and the stroke endings are specially designed.

The shapes of the glyph images for the two styles are specified (with the exception of the Euro sign glyph) by reference drawings. The constant-strokewidth style drawings show the centrelines of the strokes and the contours of the ends of the strokes; the letterpress style drawings show the complete outlines of the images (see clause 14). The general dimensions and the centrelines of the images are the same for both styles.

7   OCR-B sizes

Three sizes are specified for OCR-B glyph images in order to provide for use with a wide range of printing equipment possessing differing print quality characteristics.

NOTE   The metric and inch dimensions in this European Standard are rounded and therefore consistent but not exactly equal. Either system may be used but the two should not be intermixed.

The letterpress font is specified in size I (the smallest) only. It provides the option of a variable pitch in printing as is usual with letterpress.

The constant-strokewidth font is specified in three sizes: I, III and IV.   Mechanisms using the constant-strokewidth font will usually maintain a fixed pitch.

NOTE   Size II which was defined in the original draft version of ISO 1073/II was not included in the final version.

The centrelines for the three sizes are simply related by appropriate horizontal and vertical scale factors. The factors for size III and size IV referred to size I are:

for size III: Vertical 1,333; horizontal 1,086
for size IV: Vertical 1,500; horizontal 1,500

This scale relationship does not apply to the outline shapes since nominal strokewidth is not strictly proportional to centreline dimensions.   The strokewidths for each size are shown in the reference drawings.

The glyph image with the greatest height above the base line ("A" in figure 1) in each size is DIGIT EIGHT. The image with the greatest total height is SMALL LETTER J, because of its descender.

The centreline heights of the DIGIT EIGHT are:

for size I: 2,40 mm (0,094 in)
for size III: 3,20 mm (0,126 in)
for size IV: 3,60 mm (0,142 in)

The widest glyph image in each size (except for the Alternative SMALL LETTER M) is DIGIT ZERO. Its centreline widths are:

for size I: 1,40 mm (0,055 in)
for size III: 1,52 mm (0,060 in)
for size IV: 2,10 mm (0,083 in)

In fixed-pitch printing for OCR applications, the following minimum nominal pitches are recommended:

size I: 2,54 mm (0,100 in)
size III: 2,54 mm (0,100 in)
size IV: 3,63 mm (0,143 in)

8 Typical dimensions of the nominal printed image

Typical dimensions for the nominal printed image of the letterpress font in size I are given in table 1. These dimensions are the heights above and below the horizontal base line of digits, capital and small letters, ascenders and descenders (see figure 1).

The shapes and dimensions of the constant-strokewidth glyph images are similar except that the stroke ends are rounded.

NOTE   It is recognized that some type-making and printing processes will not be able to produce sharp corners. Corners not given a specific radius should be as sharp as practicable. However, it is not necessary for OCR purposes that the radii of the corners of the nominal printed image be less than 0,08 mm (0,0035 in).

The dimensions in table 1 are for general information only. The values for individual glyph images are obtainable from the reference drawings.
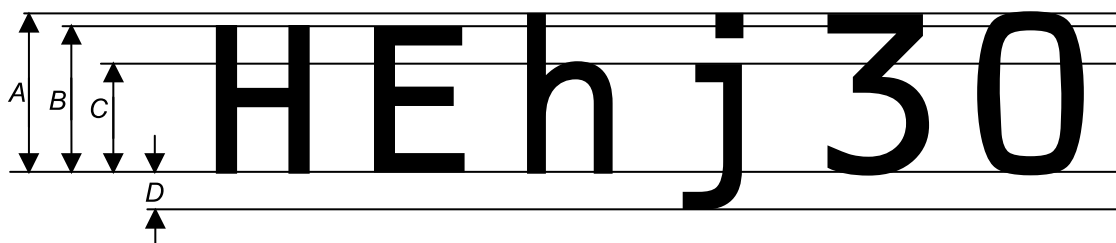
Figure 1 — Heights above and below base line

Table 1 — Typical dimensions A, B, C and D

Size	millimeters				inches			
	A	*B*	*C*	*D*	*A*	*B*	*C*	*D*
I	2,66	2,46	1,83	0,60	0,105	0,097	0,072	0,024

9   OCR-B glyph image set

The full set contains 117 glyph images and a definition for SPACE (see clause 13). Four subsets are defined in
 9.1 – 9.4 below.

9.1   Subset 1: Minimal alphanumeric subset

This subset applies to sizes I, III and IV in constant-strokewidth font and to size I in letterpress font. It contains 21
glyph images and SPACE:

0123456789

CENSTXZ

<+>| (SPACE)

9.2   Subset 2: Basic alphanumeric subset

This subset applies to sizes I and IV in constant-strokewidth font and to size I in letterpress font. It contains 25 glyph
images in addition to subset 1, i.e. a total of 46 glyph images and SPACE:

0123456789

ABCDEFGHIJKLM

NOPQRSTUVWXYZ

<+>*-=/.,| (SPACE)

9.3 Subset 3: Extended alphanumeric subset

This subset applies to sizes I and IV in constant-strokewidth font and to size I in letterpress font.

It contains 51 glyph images in addition to subset 2, i.e. 97 glyph images in all (and SPACE); in particular the images corresponding to those characters listed in ISO/IEC 646 as unique, alternative, and International Reference Version:

! " # € $ £ ¤ % & ' ( ) * + , - . /

0 1 2 3 4 5 6 7 8 9 : ; < = > ?

@ A B C D E F G H I J K L M N O

P Q R S T U V W X Y Z [\ ] ^ _

' a b c d e f g h i j k l m n o

p q r s t u v w x y z { | } ~   (SPACE)

9.4 Subset 4: Options subset

This subset applies to sizes I and IV in constant-strokewidth font and to size I in letterpress font. It contains 8 capital and 5 small national letters, 4 diacritical marks and 3 further glyph images.

Images from this subset shall be used only in conjunction with subset 3. A printing or OCR reading device may generate/recognize any of the images of this subset. The images generated/recognized by the device shall be listed in the claim of conformance (see clause 2).

Ä Å Æ IJ Ñ Ö Ø Ü   å æ ij ø ß

´ " ^

,

§ ¥ m

10 Index table

10.1 Availability of glyph images

All glyph images except the Euro sign are available in size I as constant-strokewidth font. All glyph images including the Euro sign are available in size I as letterpress font.

Only the images of the minimal alphanumeric subset (subset 1) are available in size III as constant-strokewidth font.

All images are available in size IV as constant-strokewidth font, with the exception of the Euro sign and VERTICAL LINE.

10.2 Identification of drawings

In the following table each image is given with the number of its reference drawing(s) and the subset(s) in which it is comprised.

The drawings are identified as follows :

    L: for letterpress font, size I
    C: for the constant-strokewidth font, size I
    III: for the constant-strokewidth font, size III

As stated in 14.6, the shapes for size IV are derived from those of size I for the constant-strokewidth font (designated by C).

10.3 Application considerations

Application advice is given in the column "Remarks".

It is recommended that prospective users of this European Standard consult manufacturers before deciding on a particular set of glyph images.

The ordering of the glyph images in the table does not imply any kind of coding scheme.

Table 2 — OCR-B glyph image set

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
1	1	1 L, C, III	DIGIT ONE	1 2 3	
2	2	2 L, C, III	DIGIT TWO	1 2 3	
3	3	3 L, C, III	DIGIT THREE	1 2 3	
4	4	4 L, C, III	DIGIT FOUR	1 2 3	
5	5	5 L, C, III	DIGIT FIVE	1 2 3	
6	6	6 L, C, III	DIGIT SIX	1 2 3	
7	7	7 L, C, III	DIGIT SEVEN	1 2 3	
8	8	8 L, C, III	DIGIT EIGHT	1 2 3	
9	9	9 L, C, III	DIGIT NINE	1 2 3	

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
10	O	10 L, C, III	DIGIT ZERO	1 2 3	
11	A	11 L, C	LATIN CAPITAL LETTER A	2 3	
12	B	12 L, C	LATIN CAPITAL LETTER B	2 3	
13	C	13 L, C, III	LATIN CAPITAL LETTER C	1 2 3	
14	D	14 L, C	LATIN CAPITAL LETTER D	2 3	
15	E	15 L, C, III	LATIN CAPITAL LETTER E	1 2 3	
16	F	16 L, C	LATIN CAPITAL LETTER F	2 3	
17	G	17 L, C	LATIN CAPITAL LETTER G	2 3	
18	H	18 L, C	LATIN CAPITAL LETTER H	2 3	
19	I	19 L, C	LATIN CAPITAL LETTER I	2 3	
20	J	20 L, C	LATIN CAPITAL LETTER J	2 3	
21	K	21 L, C	LATIN CAPITAL LETTER K	2 3	
22	L	22 L, C	LATIN CAPITAL LETTER L	2 3	
23	M	23 L, C	LATIN CAPITAL LETTER M	2 3	
24	N	24 L, C, III	LATIN CAPITAL LETTER N	1 2 3	
25	O	25 L, C	LATIN CAPITAL LETTER O	2 3	
26	P	26 L, C	LATIN CAPITAL LETTER P	2 3	

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
27	Q	27 L, C	LATIN CAPITAL LETTER Q	2 3	
28	R	28 L, C	LATIN CAPITAL LETTER R	2 3	
29	S	29 L, C, III	LATIN CAPITAL LETTER S	1 2 3	
30	T	30 L, C, III	LATIN CAPITAL LETTER T	1 2 3	
31	U	31 L, C	LATIN CAPITAL LETTER U	2 3	
32	V	32 L, C	LATIN CAPITAL LETTER V	2 3	
33	W	33 L, C	LATIN CAPITAL LETTER W	2 3	
34	X	34 L, C, III	LATIN CAPITAL LETTER X	1 2 3	
35	Y	35 L, C	LATIN CAPITAL LETTER Y	2 3	
36	Z	36 L, C, III	LATIN CAPITAL LETTER Z	1 2 3	
37	a	37 L, C	LATIN SMALL LETTER A	3	Smaller strokewidth; see clause 14
38	b	38 L, C	LATIN SMALL LETTER B	3	Smaller strokewidth; see clause 14
39	c	39 L, C	LATIN SMALL LETTER C	3	Smaller strokewidth; see clause 14
40	d	40 L, C	LATIN SMALL LETTER D	3	Smaller strokewidth; see clause 14
41	e	41 L, C	LATIN SMALL LETTER E	3	Smaller strokewidth; see clause 14
42	f	42 L, C	LATIN SMALL LETTER F	3	Smaller strokewidth; see clause 14
43	g	43 L, C	LATIN SMALL LETTER G	3	Smaller strokewidth; see clause 14

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
44	h	44 L, C	LATIN SMALL LETTER H	3	Smaller strokewidth; see clause 14
45	i	45 L, C	LATIN SMALL LETTER I	3	Smaller strokewidth; see clause 14
46	j	46 L, C	LATIN SMALL LETTER J	3	Smaller strokewidth; see clause 14
47	k	47 L, C	LATIN SMALL LETTER K	3	Smaller strokewidth; see clause 14
48	l	48 L, C	LATIN SMALL LETTER L	3	Smaller strokewidth; see clause 14
49	m	49 L, C	LATIN SMALL LETTER M	3	Smaller strokewidth; see clause 14
50	n	50 L, C	LATIN SMALL LETTER N	3	Smaller strokewidth; see clause 14
51	o	51 L, C	LATIN SMALL LETTER O	3	Smaller strokewidth; see clause 14
52	p	52 L, C	LATIN SMALL LETTER P	3	Smaller strokewidth; see clause 14
53	q	53 L, C	LATIN SMALL LETTER Q	3	Smaller strokewidth; see clause 14
54	r	54 L, C	LATIN SMALL LETTER R	3	Smaller strokewidth; see clause 14
55	s	55 L, C	LATIN SMALL LETTER S	3	Smaller strokewidth; see clause 14
56	t	56 L, C	LATIN SMALL LETTER T	3	Smaller strokewidth; see clause 14
57	u	57 L, C	LATIN SMALL LETTER U	3	Smaller strokewidth; see clause 14
58	v	58 L, C	LATIN SMALL LETTER V	3	Smaller strokewidth; see clause 14
59	w	59 L, C	LATIN SMALL LETTER W	3	Smaller strokewidth; see clause 14
60	x	60 L, C	LATIN SMALL LETTER X	3	Smaller strokewidth; see clause 14

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
61	y	61 L, C	LATIN SMALL LETTER Y	3	Smaller strokewidth; see clause 14
62	z	62 L, C	LATIN SMALL LETTER Z	3	Smaller strokewidth; see clause 14
63	*	63 L, C	ASTERISK	2 3	
64	+	64 L, C, III	PLUS SIGN	1 2 3	
65	▬	65 L, C	HYPHEN – MINUS	2 3	
66	=	66 L, C	EQUALS SIGN	2 3	
67	/	67 L, C	SOLIDUS	2 3	
68	▪	68 L, C	FULL STOP	2 3	
69	▟	69 L, C	COMMA	2 3	Two vertical locations are specified, one of which projects below the base line for capital letters; see 14.5 and 14.8
70	▪ ▪	70 L, C	COLON	3	
71	▪ ▟	71 L, C	SEMICOLON	3	Two vertical locations are specified, one of which projects below the base line for capital letters; see 14.5 and 14.8
72	▋▋	72 L, C	QUOTATION MARK	3	
73	▌	73 L, C	APOSTROPHE	3	
74	▬	74 L, C	LOW LINE	3	Shall be used as a stand-alone character only, and shall not be printed under another character; see clause 12
75	? ▪	75 L, C	QUESTION MARK	3	
76	! ▪	76 L, C	EXCLAMATION MARK	3	
77	(	77 L, C	LEFT PARENTHESIS	3	

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
78	)	78 L, C	RIGHT PARENTHESIS	3	
79	<	79 L, C, III	LESS-THAN SIGN	1 2 3	
80	>	80 L, C, III	GREATER-THAN SIGN	1 2 3	
81	[	81 L, C	LEFT SQUARE BRACKET	3	
82	]	82 L, C	RIGHT SQUARE BRACKET	3	
83	%	83 L, C	PERCENT SIGN	3	Smaller strokewidth; see clause 14
84	#	84 L, C	NUMBER SIGN	3	Smaller strokewidth; see clause 14
85	&	85 L, C	AMPERSAND	3	
86	@	86 L, C	COMMERCIAL AT	3	Smaller strokewidth; see clause 14
87	∧	87 L, C	UP ARROWHEAD	3	
88	¤	88 L, C	CURRENCY SIGN	3	
89	£	89 L, C	POUND SIGN	3	
90	$	90 L, C	DOLLAR SIGN	3	
91	|	91 L, C, III	VERTICAL LINE	1 2 3	
92					The glyph defined in this position in ISO 1073/II (LONG VERTICAL MARK) has been unified with the preceding one (VERTICAL LINE)
93	\	93 L, C	REVERSE SOLIDUS	3	
94	Ä	94 L, C	LATIN CAPITAL LETTER A WITH DIAERESIS	4	

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
95	Å	95 L, C	LATIN CAPITAL LETTER A WITH RING ABOVE	4	
96	Æ	96 L, C	LATIN CAPITAL LETTER AE	4	To be used also for capital ligature AE
97	Ö	97 L, C	LATIN CAPITAL LETTER O WITH DIAERESIS	4	
98	Ø	98 L, C	LATIN CAPITAL LETTER O WITH STROKE	4	
99	Ü	99 L, C	LATIN CAPITAL LETTER U WITH DIAERESIS	4	
100	IJ	100 L, C	LATIN CAPITAL LIGATURE IJ	4	
101	Ñ	101 L, C	LATIN CAPITAL LETTER N WITH TILDE	4	
102	å	102 L, C	LATIN SMALL LETTER A WITH RING ABOVE	4	Smaller strokewidth; see clause 14
103	æ	103 L, C	LATIN SMALL LETTER AE	4	To be used also for small ligature ae Smaller strokewidth; see clause 14
104	ø	104 L, C	LATIN SMALL LETTER O WITH STROKE	4	Smaller strokewidth; see clause 14
105	ij	105 L, C	LATIN SMALL LIGATURE IJ	4	Smaller strokewidth; see clause 14
106	ß	106 L, C	LATIN SMALL LETTER SHARP S	4	Smaller strokewidth; see clause 14
107	¨	107 L, C	DIAERESIS	4	For use see clause 11
108	´	108 L, C	ACUTE ACCENT	4	For use see clause 11
109	`	109 L, C	GRAVE ACCENT	3	For use see clause 11
110	^	110 L, C	CIRCUMFLEX ACCENT	4	For use see clause 11
111	~	111 L, C	TILDE	3	For use see clause 11

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
112	،	112 L, C	CEDILLA	4	For use see clause 11
113	{	113 L, C	LEFT CURLY BRACKET	3	
114	}	114 L, C	RIGHT CURLY BRACKET	3	
115	m	115 L, C	Alternative LATIN SMALL LETTER M	4	May be used in variable-pitch printing as a substitute for Ref. No. 49
116					The glyph defined in this position in ISO 1073/II (CONTINUOUS UNDERLINE) not included
117		No draw-ing	SPACE	1 2 3	SPACE is non-printing. For definition see clause 13. Not all OCR readers will necessarily recognize SPACE.
118	§	118 L, C	SECTION SIGN	4	
119	¥	119 L, C	YEN SIGN	4	
120					The glyph defined in this position in ISO 1073/II (CHARACTER ERASE) not included
121					The glyph defined in this position in ISO 1073/II (GROUP ERASE) not included
122	€	L	EURO SIGN	3	No reference drawing for this glyph image; see Annex A for shape definition

11 Use of diacritical marks

11.1 Diacritical mark repertoire

A number of diacritical marks are provided which have been designed and positioned in such a way that they can be combined with small letters to create special national-language letters. The marks are:

DIAERESIS
ACUTE ACCENT
GRAVE ACCENT
CIRCUMFLEX ACCENT
TILDE
CEDILLA

These diacritical marks may also be used free-standing.

11.2 Composite glyph images

The method for printing a composite glyph image, as well as the principle for deciding the coding of a recognized composite image, will be application-dependent, and is outside the scope of this European Standard. In particular, the naming of the diacritical marks in this standard, although corresponding to the names for free-standing marks in ISO/IEC 10646), does not imply any specific method for combination.

NOTE 1   For printing applications, a letter glyph could be followed by a backspace operation followed by a diacritical mark glyph; or the opposite sequence could be output; or the diacritical mark could be made non-spacing and followed by the letter; or the combined glyph could be composed in the printing mechanism and output in a single operation. Likewise, for OCR applications a composite glyph image recognized in a reading operation could be represented in the system as a single code assigned to the corresponding character; as a "composite graphic character" (ISO/IEC 646); as an "accented letter" (ISO/IEC 6937); as a "composite sequence" (ISO/IEC 10646-1); or by any other method defined in standards and/or specified for the particular application.

NOTE 2   A national-language letter created by combining a diacritical mark with one of the letters a-z is not necessarily a variant of that specific letter. For instance, in the scripts of some languages the character ä is a separate letter of the alphabet, distinct from and unrelated to the letter a.

The relative position of the diacritical mark and of the letter is obtained by superimposing the horizontal and vertical axes of the two glyph images concerned. Examples:

ä ç é
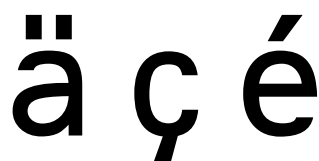
Figure 2 — Examples of composite glyph images

For OCR purposes, the superposition of a diacritical mark on the image of a basic letter must be done very accurately. Prospective users should consult manufacturers before planning inclusion of such letters in OCR glyph image sets.

11.3 Rules for glyph image combinations

Subset 4 glyph images shall not be combined with diacritical marks to produce composite images. For the subset 3 glyph images the following restrictions apply:

a)   SMALL LETTER I may be combined only with CEDILLA

b)   SMALL LETTER J shall not be combined with any diacritical mark

c)   SMALL LETTER G and SMALL LETTER Y shall not be combined with CEDILLA

d)   No letter shall be combined with multiple diacritical marks

Apart from these rules, this European Standard imposes no restrictions on letter – mark combinations. Such restrictions should therefore be specified, as necessary, for any particular printing and/or OCR application based on the OCR-B glyph images; or restrictions could be implied by definition of a character set for the application, through reference to other standards or otherwise.

The validity of any specific combination of diacritical mark and letter, as well as of any free-standing diacritical mark, is application-dependent, and outside the scope of this standard.

NOTE   Not all letter-mark combinations permitted by the rules above will be valid national-language letters.

12 Use of the LOW LINE glyph

The glyph image LOW LINE shall be used in OCR applications free-standing only, and shall not be combined with (i.e. printed under) another image.

$$DH\_1925$$

Figure 3 – Example of use of LOW LINE

13 SPACE

The SPACE is an intentionally blank position in a line of printing. With constant-pitch printing, its nominal width is equal to the printing pitch (for example, 2,54 mm if the glyph images are printed 10 per 25,4 mm). With variable-pitch printing, its nominal width is equal to the largest glyph image width available.

14 Glyph image shape definitions

14.1 Reference drawings

The shapes and dimensions of the OCR-B glyphs images for both the letterpress and the constant-strokewidth fonts are specified by drawings for size I and III. The shape of the Euro sign glyph image is however an exception, defined instead according to the syntax of ISO/IEC 9541-3; see Annex A.

The glyph images are drawn at scale 100:1 on a 2 mm square grid. The total grid measures 280 mm × 380 mm. For the purpose of illustration, two of these drawings are reproduced in Annex E at reduced size.

Grid readings should be made only from drawings on stable material. Photographic reproductions of drawings printed on paper are not satisfactory for this purpose, since the dimensional stability of paper is not sufficient.

Points on the reference drawing can be determined with an accuracy of half a square (10 µm at full size), and if desired one quarter of a square (5 µm at full size) should be possible.

14.2 Availability of duplicates

Original drawings on a stable base at 100:1 scale with the 280 mm × 380 mm grid exist in the following sets:

— OD1   Letterpress font, size I.

— OD2   Letterpress font, size I with the grid removed over approximately 2 mm around the glyph image outline. This set is particularly suitable for photographic reduction.

— OD3   Constant-strokewidth font, size I.

— OD4   Constant-strokewidth font, size III.

For detail information on acquisition of drawings see Annex F.

14.3 Type dimensions

Attention is called to the fact that this standard specifies the shapes and dimensions of the nominal printed images. These dimensions are not necessarily the same as those used in the actual image generation process. Corrections for systematic effects occurring in the printing could be needed.

Details on dimensional tolerances etc. for the printed images are specified in the International Standard ISO 1831.

14.4 Constant-strokewidth font, size I

The nominal printed image of each glyph image is defined by its centreline and by its nominal strokewidth. The nominal strokewidth is:

— 0,35 mm (0,014 in) for most of the images

— 0,31 mm (0,012 in) for all small letters and the three images #, % and @.

The centreline and preferred line endings and corners are given in drawings marked "C". Pointers establish the vertical position (base line) and the orientation. Another pointer establishes the horizontal position for fixed-pitch printing.

The reference drawings for the COMMA and the SEMICOLON contain pointers to indicate two alternative vertical positions (base lines). Either position can be selected for the two glyph images for a specific font implementation, depending on the intended use of the font.

A special effort should be made in type design and manufacturing to arrive at actual print that conforms as closely as possible to the given line endings and corners. This is especially important for the square corners of capital letters B and D.

A pointer is provided to produce the most aesthetic spacing of glyph images in a line of printing. However, on printers having a significant horizontal spacing tolerance it is recommended to use the geometric centreline

of the image instead of the line defined by the pointer where necessary to achieve an acceptable image separation.

14.5 Constant-strokewidth font, size III

The nominal printed image of each glyph image is given by its centreline and by its nominal strokewidth. The nominal strokewidth is 0,38 mm (0,015 in). The 20 reference drawings for 0 1 2 3 4 5 6 7 8 9 C E N S T X Z < + > are marked "III" and include pointers.

For the size I font, image line endings and corners should be given special attention, and the spacing situation considered.

14.6 Constant-strokewidth font, size IV

The nominal printed image of each glyph image is given by its centreline and by its nominal strokewidth. The size IV centreline is derived from the corresponding size I centreline by a linear magnification of 1,5 (see 14.4). For example, an image centreline width of 2,40 mm becomes 1,5 × 2,40 mm = 3,60 mm in size IV, and so on.

The nominal strokewidth is:

— 0,50 mm (0,020 in) for most of the glyph images.

— 0,44 mm (0,017 in) for all small letters and the three images #, % and @.

Preferred line endings and corners cannot be accurately arrived at by a 1,5 magnification since the ratio of nominal strokewidths for size IV and I is not exactly 1,5. However, given a 1,5 magnification of the size I drawing, the nominal size IV constant-strokewidth image can easily be constructed.

14.7 Letterpress font, size I

The nominal printed image of each glyph image is drawn on a reference grid (see 14.1) to allow readings with any desired accuracy from drawings marked "L". Pointers establish the vertical position (base line), the orientation and the body width. A pointer also establishes the horizontal position for fixed-pitch printing.

The reference drawings for the COMMA and the SEMI-COLON contain pointers to indicate two alternative vertical positions (base lines). Either position can be selected for the two glyph images for a specific implementation, depending on the intended use of the font.

The glyph images of the letterpress font are designed with minor strokewidth variations. However, strokewidths are always close to the nominal value of 0,35 mm (0,014 in) for digits and capital letters, and of 0,31 mm (0,012 in) for small letters and the three images #, % and @.

15 Printing the letterpress and constant-strokewidth fonts

In order to print the letterpress font and to achieve the most satisfactory appearance, the printing device should be able to print sharp corners and to keep the strokewidth variations under close control. These features are not required for printing the constant-strokewidth fonts, although a special effort should be made to produce sharp corners in the capital letters B and D.

There may well be printing equipment in which the accuracy of strokewidth control is intermediate between that required in letterpress quality and that provided by, for example, high-speed printers. It is at the discretion of the manufacturers of such printing equipment to design their type so that the printed images incorporate as many as practicable of the strokewidth variations which contribute to the aesthetically satisfactory appearance of the letterpress shapes.

Care should be taken that the printed image strokes are symmetrically distributed around the centrelines as specified in this document.

16 Illustration of OCR-B

Figure 4 shows the complete OCR-B set in size I at scales 4:1 and 1:1.

0123456789

ABCDEFGHIJKLM

NOPQRSTUVWXYZ

abcdefghijklm

nopqrstuvwxyz

*+-=/.,:;"'_

?!()<>[]%#&@^

¤£$€¥§|\{} m

ÄÅÆIJÑÖØÜåæijøß

`´^"~ , (SPACE)

Figure 4 — Illustration of letterpress font size I, scales 4:1 and 1:1

Annex A
(normative)

# Definition of Euro sign glyph image (ISO/IEC 9541-3 syntax)

```
533 475 635 0 rpe
164 96 vstem -8 81 hstem 197 80 hstem 347 80 hstem 542 81 hstem

0 hmoveto
-8 75 -45 73 -105 0 rrcurveto -125 0 -60 -110 -14 -86 rrcurveto
-76 hlineto -80 vlineto 66 hlineto
-2 -30 0 -10 1 -30 rrcurveto
-65 hlineto -80 vlineto 74 hlineto
15 -100 65 -105 121 0 rrcurveto 85 0 66 60 9 88 rrcurveto
-92 hlineto
-6 -40 -24 -27 -43 0 rrcurveto -60 0 -30 50 -12 74 rrcurveto
104 hlineto 80 vlineto -113 hlineto
-2 30 0 10 2 30 rrcurveto
148 hlineto 80 vlineto -138 hlineto
10 58 25 57 65 0 rrcurveto 50 0 18 -30 4 -37 rrcurveto
closepath

endglyph

THE DEFINITION ABOVE WILL PRODUCE THE FOLLOWING GLYPH IMAGE:
```
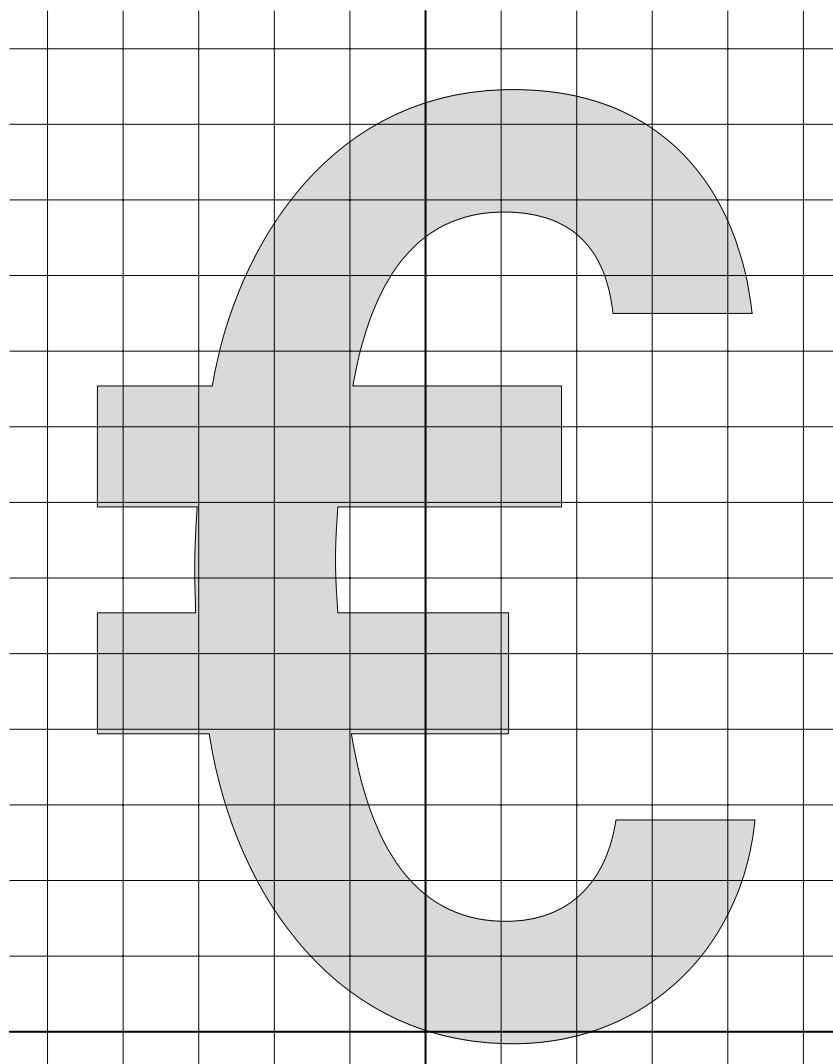
Annex B
(informative)

Main differences between ISO 1073/II-1976 and this European Standard

As compared to ISO 1073/II-1976, the following main differences exist:

a) Most occurrences of the term "character" have been replaced with the term "glyph image".

b) Original clauses "Scope" and "Field of application" have been merged as "Scope" and modified.

c) Clause "Conformance" has been added.

d) Definitions "OCR-A" and "OCR-B" have been removed and new definitions added.

e) Clause "Normative references" has been added.

f) Clause "Coding in OCR applications" has been added.

g) Clause "Style" has been rewritten.

h) In figure 1 (in clause "Typical dimensions of the nominal printed image") the letterpress font is illustrated in place of constant-strokewidth, and the text of the clause has been rewritten accordingly.

i) The text of subclauses 9.1 – 9.4 has been slightly modified. The Euro sign has been added to Subset 3. The "continuous underline" has been removed from Subset 4. Subset 5 (erase characters) has been removed.

j) The text of clause "Index table" has been rearranged and modified.

k) In table 2, the glyphs "Continuous underline" (ref. no. 116) , "Character erase" (ref. no. 120) and "Group erase" (ref. no. 121) have been removed, the "Long vertical mark" (ref. no. 92) has been unified with "Vertical line" (ref. no. 91), and the Euro sign has been added (ref. no. 122).

l) Clause "Use of diacritical marks" has been expanded for clarification.

m) Clause "Use of the two underline characters" has been modified and renamed "Use of the LOW LINE glyph" (since "Continuous underline" has been removed).

n) Clause "Character shape definition" has been renamed "Glyph image shape definition", and some textual clarifications introduced.

o) The Euro sign has been added to figure 4, the erase characters and the "Continuous underline" removed.

p) Only two of the reference drawings are now illustrated (in Annex E).

q) A number of Annexes have been added.

Annex C
(informative)

Notes on the implementation of OCR-B

The text of this annex is taken from ISO/IEC 1073/II, with minor changes Since that standard was developed at a time when OCR and printer technologies were less advanced than today, part of the text is now of limited relevance. It is however included here to provide a background to the original development of the OCR-B font.

The design of the OCR-B font is based on fundamental aesthetic principles which, as far as feasible, correspond to the criteria emerging from the long development of our classic typography. One of the essential principles prescribes that in a letter design all vertical parts must be heavier than the horizontal parts. This is also true for so-called sans serif characters, that is for a design which at first sight has a thread-like appearance. This is precisely the case for OCR-B.

The OCR-B can be implemented in two clearly different forms. It can be used as a font with constant-strokewidth as well as a letterpress font. Type generation can be based on either implementation.

For printing devices like high-speed mechanical printers and similar machines, the centreline is the skeleton along which a stroke of prescribed width is placed. For engraving, it is recommended to use a tool the diameter of which is equal to the strokewidth. The resulting engraving is completely thread-like, all strokes having an equal width. The aesthetic appearance as well as readability are partly diminished by this process.

For printing devices using other technologies, e.g. laser and ink-jet printers, the design of the font will depend on the resolution available in the device. Each font manufacturer is free to select either the constant-strokewidth or the letterpress font as a basis for the design.

The imaging resolution for some common types of printers will not permit reproduction of the finer details of the OCR-B letterpress font, like its varying strokewidth. It is however recommended that other properties of the letterpress font that can be achieved, notably the sharp angles at the end of the strokes, are implemented as far as possible. The font will then look less mechanical and bear more resemblance to the forms of traditional typography to which the human eye has been accustomed for centuries.

Annex D
(informative)

Glyph-repertoire extension needs identified in JTC 1/SC 2 revision process

For coverage of Latin-alphabet languages, two extensions to the standard were found needed:

— Changes to the standard's clause 11 "Use of diacritical marks" to permit combinations of diacritical marks not only with small letters but also with capital ones.

— Introduction of a number of special-shape letters and diacritical marks. Also introduction of a number of pre-composed letters, since composition according to clause 11.2 will in some cases result in shapes differing unacceptably from established typographical praxis.

Further, in the JTC 1/SC 2 revision process a need was identified to add those Greek capital letters that have no corresponding Latin-alphabet glyphs.

The number of special-shape and pre-composed letters to be added was however not obvious. Covering completely the needs of all Latin-alphabet languages would require the addition of an excessively large number of letters. It was therefore decided to limit the repertoire extension to letters represented in the Latin-1 to Latin-9 parts of ISO/IEC 8859.

NOTE   ISO/IEC 8859-16 (Latin-10) introduces the diacritical mark COMMA BELOW, used in two Romanian letters, in 8-bit coding. It was however not considered practical to require OCR-B discrimination of this glyph from the CEDILLA in actual applications. Instead, it was expected that the corresponding letters with cedilla could be used as a substitute (see ISO/IEC 8859-2:1998).

It was also decided that the characters added were to be included in Subset 4 (Options subset); and be available in size I letterpress font only.

This repertoire extension is illustrated in the following table, edited as a continuation of Table 2 of the standard.

The glyph image shapes used in the table are at present not formally defined. They have however been used in a tentative verification of their recognition properties.

Table D.1 — Continued OCR-B glyph image set

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
123	▬		MACRON	4	For use see clause 11
124	◡		BREVE	4	For use see clause 11
125	◢◣		CARON	4	For use see clause 11
126	″		DOUBLE ACUTE ACCENT	4	For use see clause 11
127	▪		DOT ABOVE	4	For use see clause 11
128	o		RING ABOVE	4	For use see clause 11

| Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
129	˛		OGONEK	4	For use see clause 11
130	Ą		LATIN CAPITAL LETTER A WITH OGONEK	4	
131	Đ		LATIN CAPITAL LETTER ETH	4	Shall also be used for LATIN CAPITAL LETTER D WITH STROKE
132	Ę		LATIN CAPITAL LETTER E WITH OGONEK	4	
133	Ħ		LATIN CAPITAL LETTER H WITH STROKE	4	
134	Ł		LATIN CAPITAL LETTER L WITH STROKE	4	
135	Ŧ		LATIN CAPITAL LETTER T WITH STROKE	4	
136	Þ		LATIN CAPITAL LETTER THORN	4	
137	Ŋ		LATIN CAPITAL LETTER ENG	4	
138	Œ		LATIN CAPITAL LIGATURE OE	4	
139	ą		LATIN SMALL LETTER A WITH OGONEK	4	Smaller strokewidth; see clause 14
140	đ		LATIN SMALL LETTER D WITH STROKE	4	Smaller strokewidth; see clause 14
141	ģ		LATIN SMALL LETTER G WITH CEDILLA	4	Smaller strokewidth; see clause 14
142	ħ		LATIN SMALL LETTER H WITH STROKE	4	Smaller strokewidth; see clause 14
143	ı		LATIN SMALL LETTER DOTLESS I	4	Smaller strokewidth; see clause 14

Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
144	ĵ		LATIN SMALL LETTER J WITH CIRCUMFLEX ACCENT	4	Smaller strokewidth; see clause 14
145	ł		LATIN SMALL LETTER L WITH STROKE	4	Smaller strokewidth; see clause 14
146	ŗ		LATIN SMALL LETTER R WITH CEDILLA	4	Smaller strokewidth; see clause 14
147	ŧ		LATIN SMALL LETTER T WITH STROKE	4	Smaller strokewidth; see clause 14
148	ų		LATIN SMALL LETTER U WITH OGONEK	4	Smaller strokewidth; see clause 14
149	ð		LATIN SMALL LETTER ETH	4	Smaller strokewidth; see clause 14
150	þ		LATIN SMALL LETTER THORN	4	Smaller strokewidth; see clause 14
151	ŋ		LATIN SMALL LETTER ENG	4	Smaller strokewidth; see clause 14
152	œ		LATIN SMALL LIGATURE OE	4	Smaller strokewidth; see clause 14
153	Γ		GREEK CAPITAL LETTER GAMMA	4	
154	Δ		GREEK CAPITAL LETTER DELTA	4	
155	Θ		GREEK CAPITAL LETTER THETA		
156	Λ		GREEK CAPITAL LETTER LAMDA		
157	Ξ		GREEK CAPITAL LETTER XI		
158	Π		GREEK CAPITAL LETTER PI		

| Ref. No.	Shape	Drawing(s) No.	Name	Sets	Remarks
159	Σ		GREEK CAPITAL LETTER SIGMA	4	
160	Φ		GREEK CAPITAL LETTER PHI	4	
161	Ψ		GREEK CAPITAL LETTER PSI	4	
162	Ω		GREEK CAPITAL LETTER OMEGA	4	

Annex E
(informative)

Illustration of reference drawings

The following two pages illustrate the reference drawings for the constant-strokewidth and the letterpress versions of capital letter E. The illustrations are approximately 60% full size.
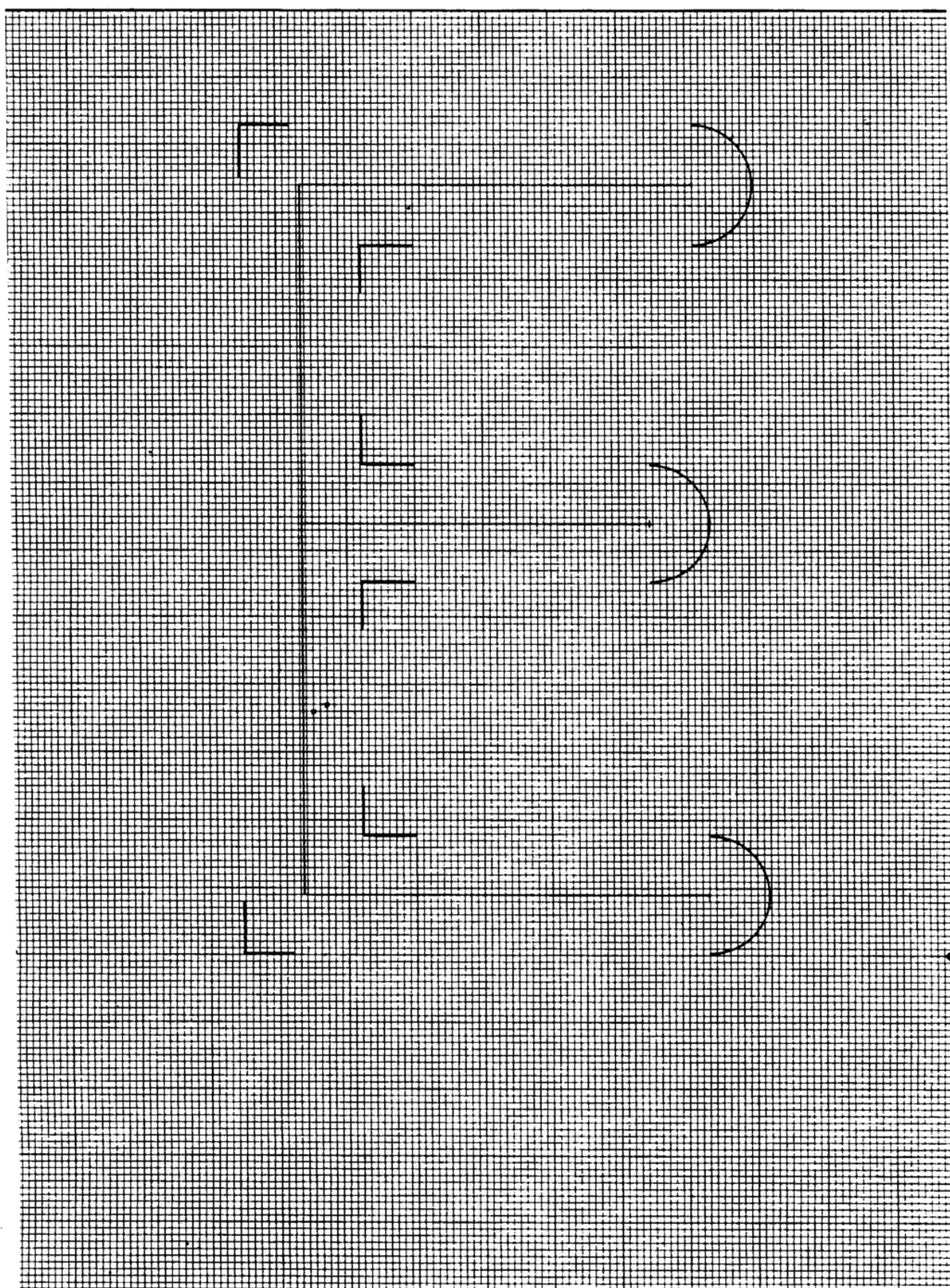
REF. DRAWING NR. 15   CAPITAL LETTER E
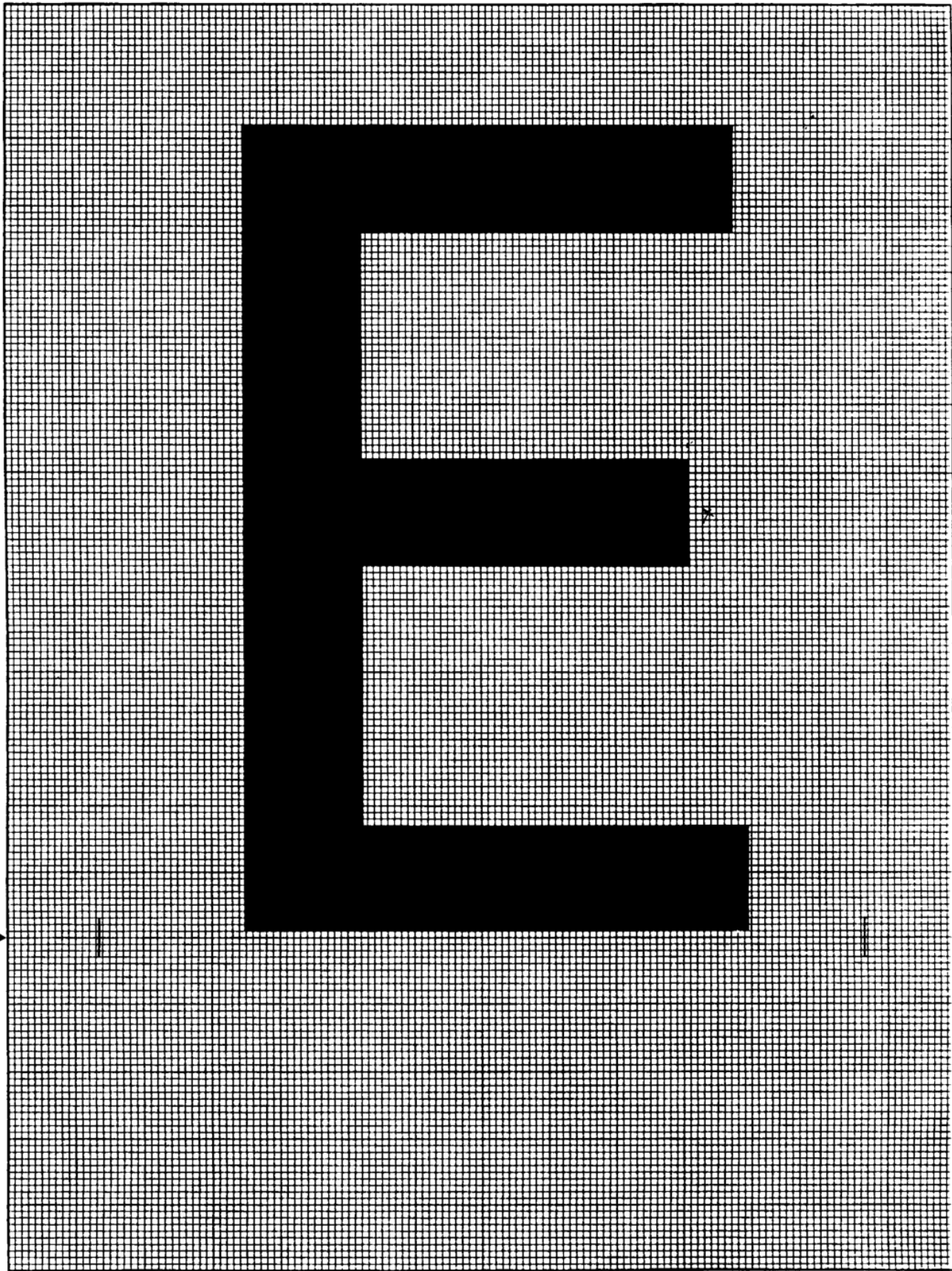
SIZE I

REF. DRAWING NR.15   CAPITAL LETTER E

SIZE I

Annex F
(informative)

Availability of reference drawings

For duplicates of the original drawings on dimensionally-stable film material contact the Secretariat of ISO/IEC JTC1/SC31:

Mr. Frank Sharkey
Uniform Code Council, Inc.
109 Lenox Drive, Suite 202
Lawrenceville, NJ 08648
USA

e-mail: sc31@uc-council.org

Reproduction and mailing costs will be charged. (Since the dimensionally-stable duplicates are comparatively expensive, prospective buyers are recommended to enquire for price before placing an order.)

Bibliography

[1]   ISO/IEC 646:1991, *Information technology — ISO 7-bit coded character set for information interchange*

[2]   ISO/IEC 6937:2001, *Information technology — Coded graphic character set for text communication — Latin alphabet*

[3]   ISO/IEC 8859, *Information technology — 8-bit single-byte coded graphic character sets* (16 parts)

[4]   ISO/IEC 10646-1:2000, *Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane*

[5]   CEN/TC304 N 837, *Draft report on the Euro glyph in OCR-B*

[6]   CEN/TC304 N 964, *CEN/TC304 project for OCR-B extension, especially Euro Sign. Interim report.*

[7]   CEN/TC304 N 982, *Report to TC304 on OCR-B situation*

BSI — British Standards Institution

BSI is the independent national body responsible for preparing
British Standards. It presents the UK view on standards in Europe and at the
international level. It is incorporated by Royal Charter.

Revisions

British Standards are updated by amendment or revision. Users of
British Standards should make sure that they possess the latest amendments or
editions.

It is the constant aim of BSI to improve the quality of our products and services.
We would be grateful if anyone finding an inaccuracy or ambiguity while using
this British Standard would inform the Secretary of the technical committee
responsible, the identity of which can be found on the inside front cover.
Tel: +44 (0)20 8996 9000. Fax: +44 (0)20 8996 7400.

BSI offers members an individual updating service called PLUS which ensures
that subscribers automatically receive the latest editions of standards.

Buying standards

Orders for all BSI, international and foreign standards publications should be
addressed to Customer Services. Tel: +44 (0)20 8996 9001.
Fax: +44 (0)20 8996 7001. Email: orders@bsi-global.com. Standards are also
available from the BSI website at http://www.bsi-global.com.

In response to orders for international standards, it is BSI policy to supply the
BSI implementation of those that have been published as British Standards,
unless otherwise requested.

Information on standards

BSI provides a wide range of information on national, European and
international standards through its Library and its Technical Help to Exporters
Service. Various BSI electronic information services are also available which give
details on all its products and services. Contact the Information Centre.
Tel: +44 (0)20 8996 7111. Fax: +44 (0)20 8996 7048. Email: info@bsi-global.com.

Subscribing members of BSI are kept up to date with standards developments
and receive substantial discounts on the purchase price of standards. For details
of these and other benefits contact Membership Administration.
Tel: +44 (0)20 8996 7002. Fax: +44 (0)20 8996 7001.
Email: membership@bsi-global.com.

Information regarding online access to British Standards via British Standards
Online can be found at http://www.bsi-global.com/bsonline.

Further information about BSI is available on the BSI website at
http://www.bsi-global.com.

Copyright