BRITISH STANDARD

BS 5464-3:
1983
ISO 2033:1983

Optical character recognition —

Part 3: Method of coding machine readable characters (MICR and OCR) for information processing

UDC 681.327.12:[681.3.04 + 003]:681.327.63

Committees responsible for this British Standard

The preparation of this British Standard was entrusted by the Office and Information Standards Committee (OIS/-) to Technical Committee OIS/2 upon which the following bodies were represented:

British Federation of Printing Machinery and Supplies Ltd.

British Radio and Electronic Equipment Manufacturers' Association

British Telecom

Business Equipment Trade Association

Computing Services Association

Independent Broadcasting Authority

Institution of Electrical Engineers

National Computing Centre Ltd.

Pira (The Research Association for the Paper and Board, Printing and Packaging Industries)

Coopted experts

This British Standard, having been prepared under the direction of the Office and Information Standards Committee, was published under the authority of the Board of BSI and comes into effect on
30 December 1983

First published as BS 4869-2 (and subsequently amended to BS 5464-3) September 1972
First revision December 1983

The following BSI references relate to the work on this standard:
Committee reference OIS/2
Draft for comment 82/61070 DC

ISBN 0 580 13510 1

Amendments issued since publication

Amd. No.	Date of issue	Comments

Contents

National foreword

This British Standard has been prepared under the direction of the Office and Information Standards Committee and is identical with ISO 2033:1983 "*Information processing — Coding of machine readable characters (MICR and OCR)*" published by the International Organization for Standardization (ISO).

This British Standard supersedes BS 5464-3, first published as BS 4869-2 in 1972 and amended in 1977, which is withdrawn.

Terminology and conventions. The text of the International Standard has been approved as suitable for publication as a British Standard without deviation. Some terminology and certain conventions are not identical with those used in British Standards; attention is drawn especially to the following.

Wherever the words "International Standard" appear, referring to this standard, they should be read as "British Standard".

The forms of expression used throughout this standard are not consistent with those normally used for British Standards.

In particular, wherever the auxiliary verb "will" is used (for example, in **5.2**: " . . . will be coded") this should be read as the auxiliary verb "shall" (for example: "shall be coded") denoting a requirement.

Cross-references

International Standard	Corresponding British Standard
ISO 1004:1977	BS 4810:1980 *Specification for print for magnetic ink character recognition* (Identical) BS 5464 *Specification for optical character recognition*
ISO 1073-I:1976	Part 1:1977 *Character set OCR-A. Shapes and dimensions of the printed image* (Identical)
ISO 1073-II:1976	Part 2:1977 *Character set OCR-B. Shapes and dimensions of the printed image* (Identical)

The Technical Committee has reviewed the provisions of ISO 646:1983 and ISO 2022:1982, to which reference is made in the text, and has decided that they are acceptable for use in conjunction with this standard.

Additional information. A revision of BS 4730 is in course of preparation to implement ISO 646:1983 for use in the UK. The first edition of ISO 2022, published in 1973, is technically equivalent to BS 4953. It has been decided to withdraw BS 4953, pending further revision of ISO 2022:1982, and it is intended to publish a new British Standard when the next edition of ISO 2022 has been finalized.

A British Standard does not purport to include all the necessary provisions of a contract. Users of British Standards are responsible for their 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, pages i and ii, pages 1 to 12, an inside back cover and a back cover.

This standard has been updated (see copyright date) and may have had amendments incorporated. This will be indicated in the amendment table on the inside front cover.

1 Scope

This International Standard defines the coded representation of printed characters recognized by reading equipment. It includes the fonts:

E 13 B	as covered in ISO 1004
CMC 7	as covered in ISO 1004
OCR-A	as covered in ISO 1073-1
OCR-B	as covered in ISO 1073-2

2 Field of application

This International Standard assigns bit-patterns to characters recognized by reading equipment. This coded information generated by the reading equipment is given to the recipient by different media, such as magnetic tape, by data transmission or a direct link. This coded representation can also be used by printing devices to print the information which shall later be read. It is not intended for general information interchange.

Two different applications are considered:

— Single-font reader:

The reading equipment is only capable of recognizing one font at a time.

— Multiple-font reader:

The reading equipment is capable of recognizing multiple fonts at the same time.

3 References

ISO 646, *Information processing — 7-bit coded character set for information interchange.*

ISO 1004, *Information processing — Magnetic ink character recognition — Print specifications.*

ISO 1073, *Alphanumeric character sets for optical recognition — Part 1: Character set OCR-A — Shapes and dimensions of the printed image — Part 2: Character set OCR-B — Shapes and dimensions of the printed image.*

ISO 2022, *Information processing — ISO 7-bit and 8-bit coded character sets — Code extension techniques.*

4 Coding

The coding given in this International Standard is based on the 7-bit code described in ISO 646 and on its extension to 8 bits according to ISO 2022.

The empty positions in code Table 1 to Table 5 are reserved for future standardization.

This International Standard does not define the character set to be read by the reading equipment.

Two codings are shown. The 8-bit coding is primarily intended for use with multi-font readers in the case where the 7-bit coding is not sufficient to represent the needed characters.

Independent of the coding shown in this International Standard, the code extension techniques given in ISO 2022 are applicable, i.e. the 7-bit coding shown in this International Standard may be transformed into 8-bit coding and the 8-bit coding shown in this International Standard may be transformed into 7-bit coding according to the rules of ISO 2022. Furthermore, the characters of columns 10 to 15 can equally be designated as a G1, G2 or G3 set.

References to positions of Table 1 to Table 5 in Table 6 to Table 9 are given by the notation “column number/row number”. The column numbers for 7-bit coding consist of one digit and those for 8-bit coding consist of two digits. The notations b_1 to b_8 refer to the 7 bits or 8 bits of the coding whereby, b_1 is the low order bit.

Example

Capital letter F is shown in position 4/6 of the 7 bit table and 04/6 of the 8-bit table. This corresponds to bit pattern 1000110 and 01000110 respectively.

4.1 7-bit coding

The 7-bit coding can be used whenever the number of characters shown is sufficient for the application. This coding can also be used within an 8-bit environment by adding an eighth bit with the value 0, as defined in ISO 2022.

4.2 8-bit coding

The 8-bit coding can be used whenever the number of characters in a 7-bit table is insufficient for the application. The 8-bit coding can also be used within a 7-bit environment as defined in ISO 2022.

Table 1 — 8-bit coding of the characters of all fonts

b_4 b_3 b_2 b_1		b_8	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
		b_7	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
		b_6	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
		b_5	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
			00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
0 0 0 0	0			SP	0	@	P		p								Å	å
0 0 0 1	1			!	1	A	Q	a	q								Æ	æ
0 0 1 0	2			"	2	B	R	b	r								Ä	ä
0 0 1 1	3			#	3	C	S	c	s			£						
0 1 0 0	4			¤	4	D	T	d	t			$						
0 1 0 1	5			%	5	E	U	e	u			¥						
0 1 1 0	6			&	6	F	V	f	v								IJ	ij
0 1 1 1	7			'	7	G	W	g	w			§						
1 0 0 0	8			(	8	H	X	h	x								Ñ	
1 0 0 1	9			)	9	I	Y	i	y								Ø	ø
1 0 1 0	10	LF	SUB	*	:	J	Z	j	z								Ö	ö
1 0 1 1	11			+	;	K	[	k	{									ß
1 1 0 0	12			,	<	L	\	l	|									
1 1 0 1	13	CR		−	=	M	]	m	}								Ü	ü
1 1 1 0	14			.	>	N	^	n										
1 1 1 1	15			/	?	O	_	o	DEL									

NOTE The empty positions are reserved for future standardization.

Table 2 — 7-bit coding of the characters of the CMC 7 font

b_7	0	0	0	0	1	1	1	1
b_6	0	0	1	1	0	0	1	1
b_5	0	1	0	1	0	1	0	1
b_4 b_3 b_2 b_1	0	1	2	3	4	5	6	7
0 0 0 0	0							
0 0 0 1	1							
0 0 1 0	2							
0 0 1 1	3							
0 1 0 0	4							
0 1 0 1	5							
0 1 1 0	6							
0 1 1 1	7							
1 0 0 0	8							
1 0 0 1	9							
1 0 1 0	10	LF	SUB					
1 0 1 1	11							
1 1 0 0	12							
1 1 0 1	13	CR						
1 1 1 0	14							
1 1 1 1	15							

NOTE The empty positions are reserved for future standardization.

Table 3 — 7-bit coding of the characters of the OCR-A font

b_7	0	0	0	0	1	1	1	1
b_6	0	0	1	1	0	0	1	1
b_5	0	1	0	1	0	1	0	1
b_4 b_3 b_2 b_1	0	1	2	3	4	5	6	7
0 0 0 0	0			0		P		
0 0 0 1	1			1	A	Q		
0 0 1 0	2		"	2	B	R		
0 0 1 1	3		£	3	C	S		
0 1 0 0	4		$	4	D	T		
0 1 0 1	5		%	5	E	U		
0 1 1 0	6		&	6	F	V		
0 1 1 1	7		'	7	G	W		
1 0 0 0	8		(	8	H	X		
1 0 0 1	9		)	9	I	Y		
1 0 1 0	10	LF	SUB	*	:	J	Z	
1 0 1 1	11			+	;	K		
1 1 0 0	12			,	<	L		|
1 1 0 1	13	CR		-	=	M		
1 1 1 0	14			.	>	N		
1 1 1 1	15			/	?	O		DEL

NOTE The empty positions are reserved for future standardization.

Table 4 — 7-bit coding of the characters of the OCR-B font

b_7	0	0	0	0	1	1	1	1
b_6	0	0	1	1	0	0	1	1
b_5	0	1	0	1	0	1	0	1
b_4 b_3 b_2 b_1	0	1	2	3	4	5	6	7
0 0 0 0	0			SP	0	@	P		p
0 0 0 1	1			!	1	A	Q	a	q
0 0 1 0	2			"	2	B	R	b	r
0 0 1 1	3			#	3	C	S	c	s
0 1 0 0	4			¤	4	D	T	d	t
0 1 0 1	5			%	5	E	U	e	u
0 1 1 0	6			&	6	F	V	f	v
0 1 1 1	7			'	7	G	W	g	w
1 0 0 0	8			(	8	H	X	h	x
1 0 0 1	9			)	9	I	Y	i	y
1 0 1 0	10	LF	SUB	*	:	J	Z	j	z
1 0 1 1	11			+	;	K	[	k	{
1 1 0 0	12			,	<	L	\	l	|
1 1 0 1	13	CR		-	=	M	]	m	}
1 1 1 0	14			.	>	N	^	n	
1 1 1 1	15			/	?	O	_	o	DEL

NOTE The empty positions are reserved for future standardization.

Table 5 — 7-bit coding of the characters of the E 13 B font

b_7	0	0	0	0	1	1	1	1
b_6	0	0	1	1	0	0	1	1
b_5	0	1	0	1	0	1	0	1
b_4 b_3 b_2 b_1	0	1	2	3	4	5	6	7
0 0 0 0	0							
0 0 0 1	1							
0 0 1 0	2							
0 0 1 1	3							
0 1 0 0	4							
0 1 0 1	5							
0 1 1 0	6							
0 1 1 1	7							
1 0 0 0	8							
1 0 0 1	9							
1 0 1 0	10	LF	SUB					
1 0 1 1	11							
1 1 0 0	12							
1 1 0 1	13	CR						
1 1 1 0	14							
1 1 1 1	15							

NOTE The empty positions are reserved for future standardization.

5 General considerations

5.1 End of line

If the information read by the equipment is structured in lines and if this structure should be maintained two possibilities are given:

— if the information is handled on a basis of records, then every line will form one record;

— if the information is handled character-by-character (data stream), the end of a line will be coded by means of control characters CR and LF.

5.2 Characters in error

If a character is recognized but cannot be identified as one character of the character set, control character SUB will be coded.

6 Font CMC 7

The characters of the CMC 7 font will be assigned to the positions of the code tables as specified in Table 6 (see also Table 1 and Table 2).

Table 6 — Characters of the CMC 7 font and their assignment to positions in the code tables

Characters of CMC 7	8-bit code	7-bit code
Digits 0 to 9	03/0 to 03/9	3/0 to 3/9
Capital letters A to Z	04/1 to 05/10	4/1 to 5/10
 Symbol S I	10/10	3/10
 Symbol S II	10/11	3/11
 Symbol S III	10/12	3/12
 Symbol S IV	10/13	3/13
 Symbol S V	10/14	3/14

7 Font OCR-A

The characters of the OCR-A font will be assigned to the positions of the code tables specified in Table 7 (see also Table 1 and Table 3).

7.1 Erase characters

Erase characters will normally be ignored by the reading equipment. If there is a requirement to code these characters, control character DEL shall be coded. For the character Group Erase one or more DEL may be coded.

Table 7 — Characters of the OCR-A font and their assignment to positions in the code tables

Characters of OCR-A	8-bit code	7-bit code
Digits 0 to 9	03/0 to 03/9	3/0 to 3/9
Capital letters A to Z	04/1 to 05/10	4/1 to 5/10
Full stop (period)[a]	02/14	2/14
Comma[a]	02/12	2/12
Equals sign	03/13	3/13
 Plus sign	02/11	2/11
 Hyphen, minus sign[a]	02/13	2/13
 Solidus	02/15	2/15
 Asterisk	02/10	2/10
 Abstract symbol H1 (hook)	11/12	3/12
 Abstract symbol H2 (fork)[a]	11/13	5/13
Abstract symbol H3 (chair)	11/14	3/14
Abstract symbol H4 (long vertical mark)	07/12	7/12
 Character erase[a]	07/15	7/15
Group erase[a]	07/15	7/15
Capital letter Å	14/0	—
Capital letter Ä	14/2	—
Capital letter Æ	14/1	—
Capital letter Ñ	14/8	—
Capital letter ∅	14/9	—
Capital letter Ö	14/10	—
Capital letter Ü	14/13	—
Colon	03/10	3/10
Semicolon	03/11	3/11
Question mark[a]	03/15	3/15
Quotation mark	02/2	2/2

[a] See special description in the following clauses.

Table 7 — Characters of the OCR-A font and their assignment to positions in the code tables

Characters of OCR-A		8-bit code	7-bit code
'	Apostrophe[a]	02/7	2/7
{	Left curly bracket	07/11	2/8
}	Right curly bracket	07/13	2/9
%	Percent sign	02/5	2/5
&	Ampersand	02/6	2/6
$	Dollar sign	10/4	2/4
£	Pound sign	10/3	2/3
¥	Yen sign	10/5	—

[a] See special description in the following clauses.

7.2 Alternative characters

If there are alternative representations of one character, all of them will be assigned to the same position. This applies to characters Full stop, Comma, Hyphen, Question mark and Apostrophe.

7.3 Abstract symbol H2 (fork)

In general, this symbol is not used in conjunction with the alphabetic characters because of its potential interference with the letter Y.

However, when it is possible to recognize both characters correctly, the indicated coding shall be used.

8 Font OCR-B

The characters of the OCR-B font will be assigned to the positions of the code table as specified in Table 8 (see also Table 1 and Table 4).

Table 8 — Characters of the OCR-B font and their assignment to positions in the code tables

Characters of OCR-B (number according to ISO 1073)			8-bit code	7-bit code
1 to 10 digits 0 to 9			03/0 to 03/9	3/0 to 3/9
11 to 36 Capital letters A to Z			04/1 to 05/10	4/1 to 5/10
37 to 62 Small letters a to z			06/1 to 07/10	6/1 to 7/10
63	*	Asterisk	02/10	2/10
64	+	Plus sign	02/11	2/11
65	-	Hyphen (minus sign)	02/13	2/13
66	=	Equals sign	03/13	3/13
67	/	Solidus	02/15	2/15
68	.	Full stop (period)	02/14	2/14
69	,	Comma	02/12	2/12
70	:	Colon	03/10	3/10
71	;	Semicolon	03/11	3/11
72	"	Quotation mark	02/2	2/2
73	'	Apostrophe	02/7	2/7
74	_	Discontinuous underline[a]	05/15	5/15
75	?	Question mark	03/15	3/15
76	!	Exclamation mark	02/1	2/1
77	(	Left parenthesis	02/8	2/8
78	)	Right parenthesis	02/9	2/9
79	<	Less than sign	03/12	3/12
80	>	Greater than sign	03/14	3/14
81	[	Left square bracket	05/11	5/11
82	]	Right square bracket	05/13	5/13
83	%	Percent sign	02/5	2/5
84	#	Number sign	02/3	2/3

[a] See special description in the following clauses.

Table 8 — Characters of the OCR-B font and their assignment to positions in the code tables

Characters of OCR-B (numbers according to ISO 1073)			8-bit code	7-bit code
85	&	Ampersand	02/6	2/6
86	@	Commercial at	04/0	4/0
87	^	Upward arrow head	05/14	5/14
88	¤	Currency sign	02/4	2/4
89	£	Pound sign	10/3	—
90	$	Dollar sign	10/4	—
91	|	Vertical line[a]	07/12	7/12
92	¦	Long vertical line[a]	07/12	7/12
93	\	Revers solidus	05/12	5/12
94	Ä	Capital letter Ä	14/2	—
95	Å	Capital letter Å	14/0	—
96	Æ	Capital letter Æ	14/1	—
97	Ö	Capital letter Ö	14/10	—
98	Ø	Capital letter ∅	14/9	—
99	Ü	Capital letter Ü	14/13	—
100	IJ	Capital letter IJ	14/6	—
101	Ñ	Capital letter Ñ	14/8	—
102	ä	Small letter ä	15/0	—
103	æ	Small letter æ	15/1	—
104	ø	Small letter ∅	15/9	—
105	ij	Small letter ij	15/6	—
106	ß	Small letter ß	15/11	—
107	¨	Diaeresis[a]	—	—
107a	ä	Small letter ä	15/2	—
107b	ö	Small letter ö	15/10	—
107c	ü	Small letter ü	15/13	—

Table 8 — Characters of the OCR-B font and their assignment to positions in the code tables

Characters of OCR-B (numbers according to ISO 1073)			8-bit code	7-bit code
108	´	Acute accent[a]	—	—
109	`	Grave accent[a]	—	—
110	^	Circumflex accent[a]	—	—
111	~	Tilde[a]	—	—
112	¸	Cedilla[a]	—	—
113	{	Left curly bracket	07/11	7/11
114	}	Right curly bracket	07/13	7/13
115	m	Alternative small letter m[a]	06/13	6/13
116	—	Continuous underline[a]	—	—
117		Space[a]	02/0	2/0
118	§	Paragraph (clause sign)	10/7	—
119	¥	Yen sign	10/5	—
120	■	Character erase[a]	07/15	7/15
121	⫽	Group erase[a]	07/15	7/15

[a] See special description in the following clauses.

8.1 Underline characters

Two characters are provided for underlining

No. 74 Discontinuous underline

No. 116 Continuous underline

The latter, Continuous underline is not intended for use in OCR applications. The character Discontinuous underline shall be used in OCR applications as a free-standing character only, and shall not be printed under another character. It shall be coded as indicated.

8.2 Diacritical marks

The characters

No. 107 Diaeresis

No. 108 Acute accent

No. 109 Grave accent

No. 110 Circumflex accent

No. 111 Tilde

No. 112 Cedilla

will be used as a diacritical mark combined with a letter. In general, reading equipment recognizes such a combined character as one single character and will code it as one single character. If such a combined character is required and it is not contained in the code table as a single character, it shall be included in future revisions of this International Standard.

8.3 Space character

If the character space (No. 117) is recognized by the reading equipment, it may be coded as one or more characters SP.

8.4 Erase characters

Erase characters (No. 120 and 121) will normally be ignored by the reading equipment. If there is a requirement to code these characters, control character DEL is to be coded. For the character Group Erase (No. 121) one or more DEL may be coded.

8.5 Small letter m

The characters

No. 49 Small letter m

No. 115 Alternative small letter m

represent the same character and are coded in the same position in the code table.

8.6 Vertical line

The characters

No. 91 Vertical line

No. 92 Long vertical line

represent the same characters and are coded in the same position in the code table.

9 Font E 13 B

The characters of the E 13 B font will be assigned to the positions in the code table as specified in Table 9 (see also Table 1 and Table 5).

Table 9 — Characters of the E 13 B font and their assignment to positions in the code tables

Characters of E 13 B	8-bit code	7-bit code
Digits 0 to 9	03/0 to 03/9	3/0 to 3/9
Symbol 1 	13/10	3/10
Symbol 2 	13/11	3/11
Symbol 3 	13/12	3/12
Symbol 4 	13/13	3/13

Annex Main differences between ISO 2033:1972 and the present edition

(This annex does not form part of the standard.)

The main differences between the first and second editions of this International Standard are the following:

a) the 4-bit coding has been removed;

b) the character Vertical line of font OCR-A is now coded in position 7/12 in 7-bit to be in accordance with the IRV of ISO 646;

c) the characters of font E 13 B have been added;

d) a coding for multi-font readers has been added;

e) the alternative assignments of OCR-B characters in positions 2/3 and 2/4 have been removed to be in accordance with the IRV of ISO 646.

blank

Publications referred to

See national foreword.

BS 5464-3: 1983
ISO 2033:1983

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: 020 8996 9000. Fax: 020 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: 020 8996 9001. Fax: 020 8996 7001.

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: 020 8996 7111. Fax: 020 8996 7048.

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: 020 8996 7002. Fax: 020 8996 7001.

Copyright

BSI
389 Chiswick High Road
London
W4 4AL