Specification of the EBU
Subtitling data exchange format

Contents

Subtitling data exchange format

1. Introduction

The importance of subtitling is increasing in the television and film industries. In many European countries, subtitling is the most common method of conveying the content of foreign language dialogue to the audience, and a broadcaster's audience may now include several major linguistic groups (notably a satellite-broadcaster).

Subtitles are also provided increasingly by broadcasters to meet the needs of the significant numbers of deaf and hard-of-hearing viewers.

A broadcaster may use two methods of providing subtitles to its audience. Subtitles may be inlaid into the picture at source. This is known as in-vision or "open" subtitling. Subtitles may also be conveyed to the audience via encoded data added to the transmitted signal (e.g. teletext), also known as "closed" subtitling.

Considering that:

- a broadcaster may wish to buy or sell programmes complete with subtitles already available in an appropriate language;

- a broadcaster may use an external company for the supply of some or all its subtitles;

- broadcasters may wish to buy compatible subtitling equipment from a number of independent suppliers;

- broadcasters may wish to harmonize the storage of subtitle data for in-vision and encoded subtitles;

the EBU has standardised a data file format for use with a personal computer to enable the exchange of in-vision and teletext subtitles. The format is intended for use by broadcasters at both national and international levels and is described below.

2. Medium for exchange

The medium for exchange is a 3.5-inch high-density portable magnetic disk (microfloppy). The disk is formatted for 1.44 Mbytes (2 sides, 80 tracks, 18 sectors/track).

3. Operating system and filename

The datafile format is defined for use with an IBM PC/XT/AT or compatible computer. The format is based upon the operating system MS/PC-DOS, version 3.3. If other operating systems are used, the datafile must be readable and writeable with MS/PC-DOS, version 3.3.

The filename should be a valid MS/PC-DOS name. It should have the file extension .STL to indicate that the datafile conforms to the EBU standard subtitle data exchange format.

Table 1 - Structure of the General Subtitle Information (GSI) block.

Byte	Bytes allocated	Information	Mnemonic
0..2	3	Code Page Number	CPN
3..10	8	Disk Format Code	DFC
11	1	Display Standard Code	DSC
12..13	2	Character Code Table number	CCT
14..15	2	Language Code	LC
16..47	32	Original Programme Title	OPT
48..79	32	Original Episode Title	OET
80..111	32	Translated Programme Title	TPT
112..143	32	Translated Episode Title	TET
144..175	32	Translator's Name	TN
176..207	32	Translator's Contact Details	TCD
208..223	16	Subtitle List Reference Code	SLR
224..229	6	Creation Date	CD
230..235	6	Revision Date	RD
236..237	2	Revision number	RN
238..242	5	Total Number of Text and Timing Information (TTI) blocks	TNB
243..247	5	Total Number of Subtitles	TNS
248..250	3	Total Number of Subtitle Groups	TNG
251..252	2	Maximum Number of Displayable Characters in any text row	MNC
253..254	2	Maximum Number of Displayable Rows	MNR
255	1	Time Code: Status	TCS
256..263	8	Time Code: Start-of-Programme	TCP
264..271	8	Time Code: First In-Cue	TCF
272	1	Total Number of Disks	TND
273	1	Disk Sequence Number	DSN
274..276	3	Country of Origin	CO
277..308	32	Publisher	PUB
309..340	32	Editor's Name	EN
341..372	32	Editor's Contact Details	ECD
373..447	75	Spare Bytes	
448..1023	576	User-Defined Area	UDA

4. Structure of the datafile[1]

4.1. Basic structure of the datafile

A subtitle list consists of a set of subtitle texts and control information for programme material in one principal language.

The datafile comprises one General Subtitle Information (GSI) block and a number of Text and Timing Information (TTI) blocks. The GSI block is placed first in the datafile, and is followed by the TTI blocks (see *Fig. 1* below).
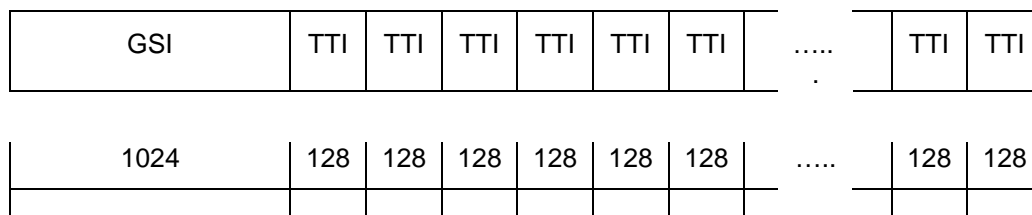
GSI	TTI	TTI	TTI	TTI	TTI	TTI	….. .	TTI	TTI
1024	128	128	128	128	128	128	…..	128	128

Fig. 1 - Basic structure of the datafile

The GSI block carries general information such as display standard, language, programme title in original and local language, etc.

A TTI block normally includes the information necessary to define one subtitle. Exceptionally, additional TTI blocks, called Extension Blocks, may be used.

The GSI and TTI blocks each include user-definable areas.

4.2. General Subtitle Information (GSI) block

4.2.1. GSI block data structure

The General Subtitle Information (GSI) block consists of 1024 bytes. The first 448 bytes are specified by the EBU, and the following 576 bytes may be defined by the user (see *Table 1*) .

The GSI block includes information:

- necessary for the use of TTI blocks;

- calculated from the TTI blocks;

- from the translator;

- about the ownership;

- defined by the user.

To ensure that the text in the GSI block can be read directly (with the DOS TYPE or PRINT command), it is assembled using characters selected from one of a restricted range of standard code pages (see *Section 4.2.2.*).

If the file consists of more than one disk, the GSI block should be copied on each disk. The Disk Sequence Number (DSN) is changed in ascending order.

---

1. Where a character code or a decimal value is described in hexadecimal notation, the suffix h is used: e.g. FFh.

4.2.2.   GSI block description

Code Page Number (CPN)

The number of the code page used in the GSI block. For international exchanges, one of the five code pages supported by MS/PC-DOS, version 3.3 must be used in the GSI block. These code pages are listed below and reproduced in *Appendix 1*.

Other code pages may be used within a given national environment (e.g.: Greek code page 928).

Code Page Number (CPN)	Character set	Hex representation		
437	United States	34h	33h	37h
850	Multilingual	38h	35h	30h
860	Portugal	38h	36h	30h
863	Canada-French	38h	36h	33h
865	Nordic	38h	36h	35h

Disk Format Code (DFC)

The subtitling data exchange format can accept data corresponding to television frame-rates of 25 and 30 frames/s. The choice is indicated using the DFC, as follows:

Disk Format Code (DFC)	Frames per second
STL25.01	25
STL30.01	30

Display Standard Code (DSC)

One of four display modes can be defined using the DSC code, as follows:

Display Standard Code (DSC)	Display standard	Hex representation
Blank	Undefined	20h
0	Open subtitling	30h
1	Level-1 teletext	31h
2	Level-2 teletext	32h

Character Code Table (CCT) number

One of five ISO Standard character code tables can be used to define the text in the Text Field (TF) of the TTI blocks. The character code table in use is indicated by the CCT number, in accordance with the table below. The complete code tables are reproduced in *Appendix 2*.

The use of the character code tables is described in *Section 5*.

Character Code Table (CCT)	Language group	ISO Standard [1,2]	Hex representation	
00	Latin	6937/2-1983/Add.1:1989	30h	30h
01	Latin/Cyrillic	8859/5-1988	30h	31h
02	Latin/Arabic	8859/6-1987	30h	32h
03	Latin/Greek	8859/7-1987	30h	33h
04	Latin/Hebrew	8859/8-1988	30h	34h

Language Code (LC)

The language for which the subtitle list is prepared is indicated by the LC code. The codes correspond to those adopted for the MAC/packet family of systems [3]. The code table is reproduced in *Appendix 3*.

Original Programme Title (OPT)

The programme title in the original language. 32 characters are available.

Original Episode Title (OET)

The title of the episode of the programme in the original language. 32 characters are available.

Translated Programme Title (TPT)

The programme title in the local language. 32 characters are available.

Translated Episode Title (TET)

The title of the episode of the programme in the local language. 32 characters are available.

Translator's Name (TN)

Name of the translator. 32 characters are available.

Translator's Contact Details (TCD)

The translator's contact details. 32 characters are available.

Subtitle List Reference Code (SLR)

This is a free-format character string which may be used to provide a unique reference for the subtitle list. 16 characters are available.

Creation Date (CD)

The date of creation of the subtitle datafile is indicated using the format defined in ISO Standard 8601 (i.e. YYMMDD).

Revision Date (RD)

The date of the most-recent modifications to the datafile is indicated using the format defined in ISO Standard 8601 (i.e. YYMMDD).

Revision Number (RN)

The revision number of the subtitle datafile may be used to specify a particular version of the subtitle list. The range is 0-99 decimal.

Total Number of TTI Blocks (TNB)

The total length of the subtitle list is given in terms of the number of TTI blocks in the datafile, including the extension blocks (if any). The range is 0-99 999 decimal. (*Note:* it is possible to store 11 242 TTI blocks on each disk.)

Total Number of Subtitles (TNS)

The number of subtitles may be equal to, or less than, the total number of TTI blocks, depending on whether or not extension blocks are used. The range is 0-99 999 decimal.

Total Number of Subtitle Groups (TNG)

The total number of subtitle groups used in the datafile. A subtitle group is defined in *Section 4.3*. The range is 0-255 decimal.

Maximum Number of Displayable Characters in any text row (MNC)

Maximum number of characters in any row of the display of the subtitles defined in the TTI blocks. The range is 0-99 decimal.

Maximum Number of Displayable Rows (MNR)

Maximum number of displayable rows per television frame which could be occupied at any one time by the subtitles defined in the TTI blocks. The range is 0-99 decimal. This parameter establishes the height of the subtitle area on the display, and can be used to ensure that all the rows of any subtitle in the list will be displayed in their entirety.

Time Code: Status (TCS)

The one-byte TCS code indicates the validity of the information given in the various GSI and TTI blocks containing time-code data (TCP and TCF in GSI block; TCI and TCO in the TTI blocks).

Time Code Status code (TCS)	Status	Hex representation
0	Not intended for use	30h
1	Intended for use	31h

Time Code: Start-of-Programme (TCP)

The time code of the first frame of the recorded video signal which is intended for transmission. The 8 bytes of the TCP code indicate, in order, the hours, minutes, seconds and frames (HHMMSSFF) of standard EBU/SMPTE time-and-control code [4].

Time Code: First in-cue (TCF)

The time code (HHMMSSFF) of the first in-cue in the subtitle list. TCF will be identical to the TCI code of the TTI block containing the first subtitle of the programme (see *Section 4.3.2.*).

Total Number of Disks (TND)

The total number of disks in the set corresponding to one complete subtitle list. The maximum number of disks is 9.

Disk Sequence Number (DSN)

The disk sequence number, starting with number 1 for the first disk in the set and increasing to the number contained in the TND code, for the last disk.

Country of Origin (CO)

The country of origin of the subtitle list is indicated using a three-letter code taken from the list in ISO Standard 3166 [5]. These codes are reproduced in *Appendix 4*.

Publisher (PUB)

Name of the publisher of the subtitle list. 32 characters are available.

Editor's Name (EN)

Name of the editor of the subtitle list. 32 characters are available.

Editor's Contact Details (ECD)

Information about the editor named in the EN code. 32 characters are available.

User-Defined Area (UDA)

The contents of this area *must not* have any effect on the reading/writing of data in any other part of the data format. This field of 576 characters may be used to carry information about the programme or subtitle list, or other relevant details.

4.2.3.   GSI block undefined values

Undefined values in the GSI block (e.g. codes 00h and 0Ah..0Fh in the TND byte) are reserved for future use.

4.2.4.   Spare bytes

The 75 spare bytes are reserved for future use.

4.2.5.   GSI block unused bytes

All unused bytes in the GSI block will be set to 20h.

4.3.   Text and Timing Information (TTI) block

4.3.1.   TTI block data structure

Each Text and Timing Information (TTI) block consists of 128 bytes. The structure of the TTI blocks is given in *Table 2* below.

A TTI block will generally contain the subtitle text together with the timing and positional data, where available, for that subtitle. TTI blocks may also contain additional subtitle data or user-specific data.

A subtitle is defined by a set of one or more TTI blocks. Each set of TTI blocks containing a given subtitle text has a unique Subtitle Number (SN). Each TTI block of a subtitle set has a unique Extension Block Number (EBN).

One or more subtitles may be grouped together, for example to establish a distinction between subtitles relating to different parts of a single programme. The group is identified by its Subtitle Group Number (SGN).

Where TTI extension blocks are used, only the first TTI block of the subtitle carries relevant information in bytes 4-15 (CS, TCI, TCO, VP, JC and CF).

TTI blocks are stored on the disk in order of continuous ascending Subtitle Numbers (SN) and ascending Time Code In (TCI) values.

Where TTI extension blocks used, they are stored on the disk in order of ascending Extension Block Number (see *Fig. 2*).

Table 2 - Structure of the Text and Timing Information (TTI) block.

Byte	Bytes allocated	Information	Mnemonic
0	1	Subtitle Group Number	SGN
1..2	2	Subtitle Number	SN
3	1	Extension Block Number	EBN
4	1	Cumulative Status	CS
5..8	4	Time Code In	TCI
9..12	4	Time Code Out	TCO
13	1	Vertical Position	VP
14	1	Justification Code	JC
15	1	Comment Flag	CF
16..127	112	Text Field	TF

4.3.2.   TTI block description

Subtitle Group Number (SGN)

This may be used to identify a particular group of subtitles. The subtitles in a subtitle group should be stored in a continuous ascending Subtitle Number order. The subtitles in the following subtitle group will be continuously numbered, without any gap, with ascending Subtitle Numbers. The time code values of one subtitle group can be different from, or equal to, the time code values of other groups. The range of the SNG code is 00h-FFh.

Subtitle Number (SN)

The SN code is a unique numeric identification for each subtitle in the list. The range is 0000h-FFFFh.

Extension Block Number (EBN)

Each TTI block in the set comprising a single subtitle contains a unique EBN in the range 00h-FFh. This byte is used to associate a maximum of 256 TTI blocks per subtitle.

EBN codes are allocated as follows:

00h..EFh :   Used to number the Extension Blocks containing additional Text Field (subtitle) information.

F0h..FDh :   Reserved for future EBU use.

FEh :   Reserved for User Data. The contents of a TTI block bearing this EBN must not have any effect on the use of the rest of the EBU data format. The Text Field of this block may be used to carry additional information. A TTI block with this indicator may be placed anywhere in the series with the same SN (see *Fig. 2*).

FFh :   Always used to indicate the last TTI block of a subtitle set. It follows that the Extension Block Number FFh is always used in cases where a single TTI block is sufficient to convey a complete subtitle.

	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	
SN:	149	150	151	152	152	152	153	154	155	156	156	157	158	159	160	…
EBN:	FFh	FFh	FFh	00h	01h	FFh	FFh	FFh	FFh	FFh	FFh	FFh	FFh	FFh	FFh	

	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	TTI	
…	880	881	882	883	883	883	883	884	885	886	886	887	888	889	890	…
	FFh	FFh	FFh	00h	01h	FFh	FFh	FFh	FFh	FFh	FFh	FFh	FFh	FFh	FFh	

Fig. 2 - Example of Extension Block Numbering

Cumulative Status (CS)

A value in the range 00h-03h, to indicate that a subtitle is part of a cumulative set of subtitles. Cumulative subtitles are also known as •add-on" subtitles, and they allow, for example, the display of a fresh subitle before the previous one has been erased from the screen.

Defined values are given in the table below.

Cumulative Status code (CS)	Significance
00h	Subtitle not part of a cumulative set
01h	First subtitle of a cumulative set
02h	Intermediate subtitle of a cumulative set
03h	Last subtitle of a cumulative set

There may be one or more intermediate subtitles with CS code 02h (or none at all, if there are two subtitles in the cumulative set).

The following example illustrates the use of this byte.

Position of subtitle in cumulative set	SN	CS	TCI	TCO
First	n	01h	10.00.00.00	10.00.15.00
Intermediate	n + 1	02h	10.00.05.00	10.00.15.00
Last	n + 2	03h	10.00.10.00	10.00.15.00

Time Code In (TCI)

Byte numbers 5-8 of the TTI block contain an EBU/SMPTE time-and-control code value indicating the start time (•in-cue") of a subtitle. The range of values is as follows:

Byte number	Time element	Decimal code range	Hex code range
5	hours	00..23	00h..17h
6	minutes	00..59	00h..3Bh
7	seconds	00..59	00h..3Bh
8	frames	00..24*	00h..18h

*In the STL30.01 format, the range is 00..29 frames (00h..1Dh).

Time Code Out (TCO)

Byte numbers 9-12 of the TTI block contain an EBU/SMPTE time-and-control code value indicating the finish time ("out-cue") of a subtitle. The range of values is as follows:

Byte number	Time element	Decimal code range	Hex code range
9	hours	00..23	00h..17h
10	minutes	00..59	00h..3Bh
11	seconds	00..59	00h..3Bh
12	frames	00..24*	00h..18h

*In the STL30.01 format, the range is 00..29 frames (00h..1Dh).

Vertical Position (VP)

This byte defines the vertical position of the first row of the subtitle.

For *teletext* subtitles, VP contains a value in the range 1-23 decimal (01h-17h) corresponding to the teletext row number of the first subtitle row.

For *in-vision* subtitles, VP contains a value in the range 0..NN decimal, where NN is the maximum number of rows indicated in the MNR field in the GSI block (*Note*: NN cannot be greater than 99 decimal (63h)). This VP represents the number of row locations from the top of the screen to the first subtitle row.

For both *teletext* and *in-vision* subtitles, the vertical positions of subsequent rows of the subtitle are defined with the carriage-return/line-feed (CR/LF) indicator in the Text Field (TF).

Justification Code (JC)

The JC code controls the horizontal alignment of the displayed subtitle. Four codes are available, as follows.

Justification Code (JC)	Significance
00h	unchanged presentation
01h	left-justified text
02h	centred text
03h	right-justified text

If JC is set to 00h, the text and box positioning are determined by all the spaces and control characters which accompany the subtitle text in the Text Field..

If JC is set to 01h, 02h or 03h, all leading and trailing spaces in the Text Field are ignored. The overall horizontal alignment of the subtitle display area will be determined by the users' equipment, and the text will be left or right justified, or centred, within that display area according to the content of the JC code.

Comment Flag (CF)

The Comment Flag is used to indicate TTI blocks which contain texts such as translator's comments, instead of subtitle data. CF may have values in the range 00h-01h, attributed as follows.

Comment Flag (CF)	Significance
00h	TF contains subtitle data
01h	TF contains comments not intended for transmission

Text Field (TF)

The Text Field contains all, or some of, the text and control characters to convey a subtitle. It has a fixed length of 112 bytes and may contain any valid character code selected from the character code table specified in the CCT field of the GSI block.

Regardless of which character code table is in use, the following conventions must be applied:

- the CR/LF indicator, used to initiate the second and subsequent rows of the subtitle display, is conveyed by character code 8Ah;

- the Text Field of the last TTI block of a subtitle must always terminate with code 8Fh;

- unused space in the Text Field will be set to 8Fh.

4.3.3.   Undefined values in the TTI block

All undefined values in the TTI block are reserved for future use.

5.   Character code tables

The character codes of the subtitle text (i.e. the Text Field information) are obtained from the character code table specified by the CCT number in the GSI block.

The character code tables to be used are reproduced in *Appendix 2*.

The accented letters in the Latin-based languages are created according to the •floating accent" principle. Column •C" of character code table 00 (Latin alphabet) contains diacritical marks which may be associated with another character addressed to the same presentation position. Each accented character occupies two bytes, and the diacritical mark is sent first (e.g. Ä = C8h 41h, ê = C3h 65h).

For *teletext* subtitles, the control codes used in the CCIR Teletext System B (fixed-format) are to be used. These codes, which occupy positions corresponding to columns 0 and 1 of the character code tables, are defined in [6] and reproduced at the end of *Appendix 2*. These codes are not intended to be used to describe in-vision subtitles.

For *in-vision* subtitles, three basic character control parameters are included: italics, underline and boxing. These codes occupy positions corresponding to the first six codes of column 8 of the character code tables (80h..85h).

The following table lists the applications and effects of all TF codes in the range 00h..FFh.

Control codes (hex)	Open or closed subtitles	Effect	Notes
00h..1Fh	Closed	Teletext control codes	See *Appendix 2*
20h..7Fh	Both	Character codes	See *Appendix 2*
80h	Open	Italics ON	-
81h	Open	Italics OFF	-
82h	Open	Underline ON	-
83h	Open	Underline OFF	-
84h	Open	Boxing ON	-
85h	Open	Boxing OFF	-
86h..89h	-	Reserved for future use	-
8Ah	Both	CR/LF	-
8Bh..8Eh	-	Reserved for future use	-
8Fh	Both	Code for unused space in TF	-
90h..9Fh	-	Reserved for future use	-
A1h..FFh	Both	Character codes	See *Appendix 2*

Appendix 1
Character code pages used in the GSI block

This Appendix reproduces the five code pages accepted for used in the General Subtitle Information (GSI) block, in the context of international subtitle exchanges. These are the five code pages supported by MS-DOS/PC-DOS, version 3.3.

Second byte ↓	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0		►		0	@	P	`	p	Ç	É	á	▒	└	╨	α	≡
1	☺	◄	!	1	A	Q	a	q	ü	æ	í	▓	┴	╤	β	±
2	☻	↕	"	2	B	R	b	r	é	Æ	ó	█	┬	╥	Γ	≥
3	♥	‼	#	3	C	S	c	s	â	ô	ú	│	├	╙	π	≤
4	♦	¶	$	4	D	T	d	t	ä	ö	ñ	┤	─	╘	Σ	⌠
5	♣	§	%	5	E	U	e	u	à	ò	Ñ	╡	┼	╒	σ	⌡
6	♠	▬	&	6	F	V	f	v	å	û	ª	╢	╞	╓	μ	÷
7	·	↨	'	7	G	W	g	w	ç	ù	º	╖	╟	╫	τ	≈
8	◘	↑	(	8	H	X	h	x	ê	ÿ	¿	╕	╚	╪	Φ	°
9	○	↓	)	9	I	Y	i	y	ë	Ö	⌐	╣	╔	┘	Θ	∙
A	◙	→	*	:	J	Z	j	z	è	Ü	¬	║	╩	┌	Ω	•
B	♂	←	+	;	K	[	k	{	ï	¢	½	╗	╦	█	δ	√
C	♀		,	<	L	\	l	|	î	£	¼	╝	╠	▄	∞	ⁿ
D	♪	↔	-	=	M	]	m	}	ì	¥	¡	╜	═	▌	ø	²
E	♫	▲	.	>	N	^	n	~	Ä	Pts	«	╛	╬	▐	ε	■
F	☼	▼	/	?	O	_	o	△	Å	ƒ	»	┐	╧	▀	∩	

Code page 437 (United states)

Second byte ↓ / First byte →

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0		►		0	@	P	`	p	Ç	É	á	░	└	õ	Ó	-
1	☺	◄	!	1	A	Q	a	q	ü	æ	í	▒	┴	Ð	ß	±
2	●	↕	"	2	B	R	b	r	é	Æ	ó	▓	┬	Ê	Ô	=
3	♥	‼	#	3	C	S	c	s	â	ô	ú	│	├	Ë	Ò	¾
4	♦	¶	$	4	D	T	d	t	ä	ö	ñ	┤	─	È	õ	¶
5	♣	§	%	5	E	U	e	u	à	ò	Ñ	Á	┼		õ	§
6	♠		&	6	F	V	f	v	å	û	ª	Â	ã	Í	µ	÷
7	·	↕	'	7	G	W	g	w	ç	ù	º	À	Ã	Î	Þ	¸
8	◘	↑	(	8	H	X	h	x	ê	ÿ	¿	©	╚	Ï	þ	°
9	○	↓	)	9	I	Y	i	y	ë	Ö	®	╣	╔	┘	Ú	¨
A	◙	→	*	:	J	Z	j	z	è	Ü	¬	║	╩	┌	Û	·
B	♂	←	+	;	K	[	k	{	ï	ø	½	╗	╦	█	Ù	¹
C	♀		,	<	L	\	l	|	î	£	¼	╝	╠	▄	ý	³
D	♪	↔	-	=	M	]	m	}	ì	Ø	¡	¢	═	¦	Ý	²
E	♫	▲	.	>	N	^	n	~	Ä	×	«	¥	╬	Ì	¯	■
F	☼	▼	/	?	O	_	o	⌂	Å	ƒ	»	┐	¤	■	´	

Code page 850 (Multilingual)

Second byte ↓ / First byte →

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0		►		0	@	P	`	p	Ç	É	á	░	└	╨	α	≡
1	☺	◄	!	1	A	Q	a	q	ü	À	í	▒	┴	╥	β	±
2	●	↕	"	2	B	R	b	r	é	È	ó	▓	┬	╤	Γ	≥
3	♥	‼	#	3	C	S	c	s	â	ô	ú	│	├	╙	π	≤
4	♦	¶	$	4	D	T	d	t	ã	õ	ñ	┤	─	╘	Σ	⌠
5	♣	§	%	5	E	U	e	u	à	ò	Ñ	╡	┼	╒	σ	⌡
6	♠	▬	&	6	F	V	f	v	Á	Ú	ª	╢	╞	╓	µ	÷
7	·	↕	'	7	G	W	g	w	ç	ù	º	╖	╟	╫	τ	≈
8	◘	↑	(	8	H	X	h	x	ê	Í	¿	╕	╚	╪	Φ	°
9	○	↓	)	9	I	Y	i	y	Ê	Õ	Ò	╣	╔	┘	Θ	·
A	◙	→	*	:	J	Z	j	z	è	Ü	¬	║	╩	┌	Ω	·
B	♂	←	+	;	K	[	k	{	Ì	¢	½	╗	╦	█	δ	√
C	♀		,	<	L	\	l	|	Ô	£	¼	╝	╠	▄	∞	ⁿ
D	♪	↔	-	=	M	]	m	}	ì	Ù	¡	╜	═	█	ø	²
E	♫	▲	.	>	N	^	n	~	Ã	Pts	«	╛	╬	█	ε	■
F	☼	▼	/	?	O	_	o	⌂	Â	Ó	»	┐	╧	▀	∩	

Code page 860 (Portugal)

Second byte ↓	First byte → 0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0		►		0	@	P	`	p	Ç	É	¦	▒	└	╨	α	≡
1	☺	◄	!	1	A	Q	a	q	ü	È	´	▒	┴	╤	β	±
2	☻	↕	"	2	B	R	b	r	é	Ê	ó	▓	┬	╥	Γ	≥
3	♥	‼	#	3	C	S	c	s	â	ô	ú	│	├	╙	π	≤
4	♦	¶	$	4	D	T	d	t	Â	Ë	¨	┤	─	╘	Σ	⌠
5	♣	§	%	5	E	U	e	u	à	Ï	¸	╡	┼	╒	σ	⌡
6	♠	▬	&	6	F	V	f	v	¶	û	³	╢	╞	╓	µ	÷
7	·	↨	'	7	G	W	g	w	ç	ú		╖	╟	╫	τ	≈
8	▫	↑	(	8	H	X	h	x	ê	¤	Î	╕	╚	╪	Φ	°
9	○	↓	)	9	I	Y	i	y	ë	Ô	⌐	╣	╔	┘	Θ	•
A	◙	→	*	:	J	Z	j	z	è	Ü	¬	║	╩	┌	Ω	·
B	♂	←	+	;	K	[	k	{	ï	¢	½	╗	╦	█	δ	√
C	♀		,	<	L	\	l	|	î	£	¼	╝	╠	▄	∞	ⁿ
D	♪	↔	-	=	M	]	m	}	=	Ù	¾	╜	═	▌	ø	²
E	♫	▲	.	>	N	^	n	~	Ä	Û	«	╛	╬	▐	ε	■
F	☼	▼	/	?	O	_	o	△	§	ƒ	»	┐	╧	▀	∩	

Code page 863 (Canada-French)

Second byte ↓	First byte → 0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0		►		0	@	P	`	p	Ç	É	á	▒	└	╨	α	≡
1	☺	◄	!	1	A	Q	a	q	ü	æ	í	▒	┴	╤	β	±
2	☻	↕	"	2	B	R	b	r	é	Æ	ó	▓	┬	╥	Γ	≥
3	♥	‼	#	3	C	S	c	s	â	ô	ú	│	├	╙	π	≤
4	♦	¶	$	4	D	T	d	t	ä	ö	ñ	┤	─	╘	Σ	⌠
5	♣	§	%	5	E	U	e	u	à	ò	Ñ	╡	┼	╒	σ	⌡
6	♠	▬	&	6	F	V	f	v	å	û	ª	╢	╞	╓	µ	÷
7	·	↨	'	7	G	W	g	w	ç	ú	º	╖	╟	╫	τ	≈
8	▫	↑	(	8	H	X	h	x	ê	ÿ	¿	╕	╚	╪	Φ	□
9	○	↓	)	9	I	Y	i	y	ë	Ö	⌐	╣	╔	┘	Θ	•
A	◙	→	*	:	J	Z	j	z	è	Ü	¬	║	╩	┌	Ω	·
B	♂	←	+	;	K	[	k	{	ï	ø	½	╗	╦	█	δ	√
C	♀		,	<	L	\	l	|	î	£	¼	╝	╠	▄	∞	ⁿ
D	♪	↔	-	=	M	]	m	}	ì	Ø	¡	╜	═	▌	ø	²
E	♫	▲	.	>	N	^	n	~	Ä	Pts	«	╛	╬	▐	ε	■
F	☼	▼	/	?	O	_	o	△	Å	ƒ	¤	┐	╧	▀	∩	

Code page 865 (Norway)

Appendix 2

Character code tables used in TTI blocks

First byte →

Second byte ↓

2nd\1st	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0			SP	0	@	P	`	p			NBSP	°		—	Ω	κ
1			!	1	A	Q	a	q			¡	±	`	¹	Æ	æ
2			"	2	B	R	b	r			¢	²	´	®	Đ	đ
3			#	3	C	S	c	s			£	³	^	©	ạ	ð
4			¤	4	D	T	d	t			$	×	~	™	Ħ	ħ
5			%	5	E	U	e	u			¥	µ	¯	♪		⌐
6			&	6	F	V	f	v				¶	˘	¬	IJ	ij
7			'	7	G	W	g	w			§	•	˙	¦	Ŀ	ŀ
8			(	8	H	X	h	x				÷	¨		Ł	ł
9			)	9	I	Y	i	y			`	′			Ø	ø
A			*	:	J	Z	j	z			"	″	°		Œ	œ
B			+	;	K	[	k	{			«	»	¸		ọ	ß
C			,	<	L	\	l	|			←	¼	_	⅛	Þ	þ
D			-	=	M	]	m	}			↑	½	″	⅜	Ŧ	ŧ
E			.	>	N	^	n	~			→	¾	˛	⅝	Ŋ	ŋ
F			/	?	O	_	o				↓	¿	ˇ	⅞	’n	SHY

Notes:

(1) The SPACE character is located on position 20_h of the code table

(2) NBSP = no-break space

(3) SHY = soft hyphen

Character code table 00 - Latin alphabet
(from ISO 6937/2-1983, Addendum 1-1989)

First byte →

Second byte ↓	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0			SP	0	@	P		p			NBSP	А	Р	а	р	N°
1			!	1	A	Q	a	q			Ё	Б	С	б	с	ё
2			"	2	B	R	b	r			Ђ	В	Т	в	т	ђ
3			#	3	C	S	c	s			Ѓ	Г	У	г	у	ѓ
4			$	4	D	T	d	t			Є	Д	Ф	д	ф	є
5			%	5	E	U	e	u			Ѕ	Е	Х	е	х	ѕ
6			&	6	F	V	f	v			І	Ж	Ц	ж	ц	і
7			'	7	G	W	g	w			Ї	З	Ч	з	ч	ї
8			(	8	H	X	h	x			Ј	И	Ш	и	ш	ј
9			)	9	I	Y	i	y			Љ	Й	Щ	й	щ	љ
A			*	:	J	Z	j	z			Њ	К	Ъ	к	ъ	њ
B			+	;	K	[	k	{			Ћ	Л	Ы	л	ы	ћ
C			,	<	L	\	l	¦			Ќ	М	Ь	м	ь	ќ
D			-	=	M	]	m	}			SHY	Н	Э	н	э	§
E			.	>	N	^	n	~			Ў	О	Ю	о	ю	ў
F			/	?	O	_	o				Џ	П	Я	п	я	џ

Note: For the Ruthenian language, the characters in code positions $A_h/5_h$ (S) and $F_h/5_h$ (s) are replaced by Ґ and ґ, respectively.

Character code table 01 - Latin/Cyrillic alphabet
(from ISO 8859/5-1988)

First byte →

Second byte ↓	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0			SP	0 •	@	P		p			NBSP			ذ	‑	ـ
1			!	1 ١	A	Q	a	q					ء	ر	ف	ّ
2			"	2 ٢	B	R	b	r					آ	ز	ق	°
3			#	3 ٣	C	S	c	s					أ	س	ك	
4			$	4 ٤	D	T	d	t			¤		ؤ	ش	ل	
5			%	5 ٥	E	U	e	u					إ	ص	م	
6			&	6 ٦	F	V	f	v					ئ	ض	ن	
7			'	7 ٧	G	W	g	w					ا	ط	ه	
8			(	8 ٨	H	X	h	x					ب	ظ	و	
9			)	9 ٩	I	Y	i	y					ة	ع	ى	
A			*	:	J	Z	j	z					ت	غ	ي	
B			+	;	K	[	k	{				؛	ث		ً	
C			,	<	L	\	l	¦			،		ج		ٌ	
D			-	=	M	]	M	}			SHY		ح		ٍ	
E			.	>	N	^	N	~					خ		َ	
F			/	?	O	_	o					؟	د		ُ	

Character code table 02 - Latin/Arabic alphabet
(from ISO 8859/6-1987)

First byte →

Second byte	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0			SP	0	@	P	`	p			NBSP	°	῏ι	Π	ῠ	π
1			!	1	A	Q	a	q			□	±	Α	Ρ	α	ρ
2			"	2	B	R	b	r			□	²	Β		β	ς
3			#	3	C	S	c	s			£	³	Γ	Σ	γ	σ
4			$	4	D	T	d	t				΄	Δ	Τ	δ	τ
5			%	5	E	U	e	u				῁	Ε	Υ	ε	υ
6			&	6	F	V	f	v			¦	Ά	Ζ	Φ	ζ	φ
7			'	7	G	W	g	w			§	·	Η	Χ	η	χ
8			(	8	H	X	h	x			¨	Έ	Θ	Ψ	θ	ψ
9			)	9	I	Y	i	y			©	Ή	Ι	Ω	ι	ω
A			*	:	J	Z	j	z				Ί	Κ	Ϊ	κ	ϊ
B			+	;	K	[	k	{			«	»	Λ	Ϋ	λ	ϋ
C			,	<	L	\	l	|			¬	Ό	Μ	ά	μ	ό
D			-	=	M	]	m	}			SHY	½	Ν	έ	ν	ύ
E			.	>	N	^	n	~				Ύ	Ξ	ή	ξ	ώ
F			/	?	O	_	o				-	Ώ	Ο	ί	o	

Character code table 03 - Latin/Greek alphabet
From ISO 8859/7-1987

First byte →

Second byte ↓

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0			SP	0	@	P		p			NBSP	°			א	נ
1			!	1	A	Q	a	q				±			ב	ס
2			"	2	B	R	b	r			¢	²			ג	ע
3			#	3	C	S	c	s			£	³			ד	ף
4			$	4	D	T	d	t			¤	´			ה	פ
5			%	5	E	U	e	u			¥	µ			ו	ץ
6			&	6	F	V	f	v			¦	¶			ז	צ
7			'	7	G	W	g	w			§	·			ח	ק
8			(	8	H	X	h	x			¨	¸			ט	ר
9			)	9	I	Y	i	y			©	¹			י	ש
A			*	:	J	Z	j	z			×	÷			ך	ת
B			+	;	K	[	k	{			«	»			כ	
C			,	<	L	\	l	|			¬	¼			ל	
D			-	=	M	]	m	}			SHY	½			ם	
E			.	>	N	^	n	~			®	¾			מ	
F			/	?	O	_	o				‾			‗	ן	

Character code table 04 - Latin/Hebrew alphabet
From ISO 8859/8-1988

Second   First byte →

Byte   0   1

↓   0

	0	Alpha black	Mosaic black
	1	Alpha Red	Mosaic red
	2	Alpha green	Mosaic green
	3	Alpha yellow	Mosaic yellow
	4	Alpha blue	Mosaic blue
	5	Alpha magenta	Mosaic magenta
	6	Alpha cyn	Mosaic cyan
	7	Alpha white	Mosaic white
	8	Flash	Conceal (2)
	9	Steady (1,2)	Contiguous mosaic (1,2)
	A	End box (1,2)	Separated mosaic (2)
	B	Start box	(5)
	C	Normal height	Black background (1,2)
	D	Double hieght	New background (2)
	E	Double width	Hold mosaic (2)
	F	Double size	Release Mosaic (1)

Notes

(1) Presumed at the start of each display row

(2) Action "set at" other are "set after"

(3) Two consecutive codes are sent, actio takes place between them

(4) No action at level 1

(5) Reserved code

Control characters for teletext subtitles

(reproduced from EBU document Tech. 3240)

Appendix 3
Code table for languages used in TTI blocks

This Appendix lists the codes used in the Language Code (LC) field of the GSI block. These codes indicate the language used in the Text Field of the Text and Timing Information (TTI) blocks. The codes are identical to those used in the systems of the MAC/packet family, and are reproduced from the EBU specification for these systems [3].

European languages written in Latin-based alphabets

Code	Language	Code	Language	Code	Language
00	Unknown/not applicable	16	Lappish	2C	
01	Albanian	17	Latin	2D	
02	Breton	18	Latvian	2E	
03	Catalan	19	Luxembourgian	2F	
04	Croatian	1A	Lithuanian	30	
05	Welsh	1B	Hungarian	31	
06	Czech	1C	Maltese	32	
07	Danish	1D	Dutch	33	
08	German	1E	Norwegian	34	
09	English	1F	Occitan	35	
0A	Spanish	20	Polish	36	
0B	Esperanto	21	Portugese	37	Reserved for
0C	Estonian	22	Romanian	38	national assignment
0D	Basque	23	Romansh	39	
0E	Faroese	24	Serbian	3A	
0F	French	25	Slovak	3B	
10	Frisian	26	Slovenian	3C	
11	Irish	27	Finnish	3D	
12	Gaelic	28	Swedish	3E	
13	Galician	29	Turkish	3F	
14	Icelandic	2A	Flemish		
15	Italian	2B	Wallon		

Other languages

Code	Language	Code	Language	Code	Language
7F	Amharic	69	Japanese	53	Shona
7E	Arabic	68	Kannada	52	Sinhalese
7D	Armenian	67	Kazakh	51	Somali
7C	Assamese	66	Khmer	50	Sranan Tongo
7B	Azerbaijani	65	Korean	4F	Swahili
7A	Bambora	64	Laotian	4E	Tadzhik
79	Bielorussian	63	Macedonian	4D	Tamil
78	Bengali	62	Malagasay	4C	Tatar
77	Bulgarian	61	Malaysian	4B	Telugu
76	Burmese	60	Moldavian	4A	Thai
75	Chinese	5F	Marathi	49	Ukrainian
74	Churash	5E	Ndebele	48	Urdu
73	Dari	5D	Nepali	47	Uzbek
72	Fulani	5C	Oriya	46	Vietnamese
71	Georgian	5B	Papamiento	45	Zulu
70	Greek	5A	Persian	44	
6F	Gujurati	59	Punjabi	43	
6E	Gurani	58	Pushtu	42	
6D	Hausa	57	Quechua	41	
6C	Hebrew	56	Russian	40	
6B	Hindi	55	Ruthenian		
6A	Indonesian	54	Serbo-croat		

Appendix 4
3-letter codes for use in the CO field

The three-letter (•Alpha-3") codes given in the table below are for use in the Country o f Origin (CO) field of the GSI block. The codes are those given in ISO Standard 3166-1988 [5]. The country names indicated here are not necessarily the official names of the countries or territories. The information given here does not imply, on the part of the European Broadcasting Union, any opinion regarding the political status of these countries or territories.

ISO "Alpha-3" code	Country	ISO "Alpha-3" code	Country
AAA-AAZ	private use, *see note 2*	BHR	Bahrain
ABW	Aruba	BHS	Bahamas
AFG	Afghanistan	BLZ	Belize
AGO	Angola	BMU	Bermuda
AIA	Anguilla	BOL	Bolivia
ALB	Albania	BRA	Brazil
AND	Andorra	BRB	Barbados
ANT	Netherlands Antilles	BRN	Brunei Darussalam
ARE	United Arab Emirates	BTN	Bhutan
ARG	Argentina	BUR	Burma
ASM	American Samoa	BVT	Bouvet Island
ATA	Antarctica	BWA	Botswana
ATF	French Southern Territories	BYS	Byelorussian SSR
ATG	Antigua and Barbuda	CAF	Central African Republic
ATN	See ATA	CAN	Canada
AUS	Australia	CCK	Cocos (Keeling) Islands
AUT	Austria	CHE	Switzerland
BDI	Burundi	CHL	Chile
BEL	Belgium	CHN	China
BEN	Benin	CIV	Côte d'Ivoire
BFA	Burkina Faso	CMR	Cameroon
BGD	Bangladesh	COG	Congo
BGR	Bulgaria	COK	Cook Islands

ISO • Alpha-3" code	Country	ISO • Alpha-3" code	Country
COL	Colombia	GUF	French Guinea
COM	Comoros	GUM	Guam
CPV	Cape Verde	GUY	Guyana
CRI	Costa Rica	HKG	Hong Kong
CSK	Czechoslovakia	HMD	Heard and Mc Donald Islands
(CTE)	See KIR	HND	Honduras
CUB	Cuba	HTI	Haiti
CXR	Christmas Island	HUN	Hungary
CYM	Cayman Islands	(HVO)	See BFA
CYP	Cyprus	IDN	Indonesia
DDR	German Democratic Republic	IDN	India
DEU	Germany, Federal Republic of	IOT	British Indian Ocean Territory
DJI	Djibouti	IRL	Ireland
DMA	Dominica	IRN	Iran
DNK	Denmark	IRQ	Iraq
DOM	Dominican Republic	ISL	Iceland
DZA	Algeria	ISR	Israel
ECU	Ecuador	ITA	Italy
EGY	Egypt	JAM	Jamaica
ESH	Western Sahara	JOR	Jordan
ESP	Spain	JPN	Japan
ETH	Ethiopia	(JTN)	See UMI
FIN	Finland	KEN	Kenya
FJI	Fiji	DHM	Kampuchea, Democratic
FLK	Falkland Islands (Malvinas)	KIR	Kiribati
FRA	France	KNA	Saint Kitts and Nevis
FRO	Faroe Islands	KOR	Korea, Republic of
FSM	Micronesia	KWT	Kuwait
GAB	Gabon	LAO	Lao, People's Democratic Republic of
GBR	United Kingdom	LBN	Lebanon
GHA	Ghana	LBR	Liberia
GIB	Gibraltar	LBY	Libyan Arab Jamahiriya
GIN	Guinea	LCA	Saint Lucia
GLP	Guadeloupe	LIE	Liechtenstein
GMB	Gambia	LKA	Sri Lanka
GNB	Guinea-Bissau	LSO	Lesotho
GNQ	Equatorial Guinea	LUX	Luxembourg
GRC	Greece	MAC	Macau
GRD	Grenada	MAR	Morocco
GRL	Greenland	MCO	Monaco
GTM	Guatemala	MDG	Madagascar

ISO "Alpha-3" code	Country	ISO "Alpha-3" code	Country
MDV	Maldives	PRK	Korea, Democratic People's Republic of
MEX	Mexico	PRT	Portugal
MHL	Marshall Islands	PRY	Paraguay
(MID)	See UMI	(PUS)	See UMI
MLI	Mali	PYF	French Polynesia
MLT	Malta	QAT	Qatar
MNG	Mongolia	QMA-QZZ	private use, see note 2
MNP	Nothern Mariana Islands	REU	Réunion
MOZ	Mozambique	ROM	Romania
MRT	Mauritania	RWA	Rwanda
MSR	Montserrat	SAU	Saudi Arabia
MTQ	Martinique	SDN	Sudan
MUS	Mauritius	SEN	Senegal
MWI	Malawi	SGP	Singapore
MYS	Malaysia	SHN	St. Helena
NAM	Namibia	SJM	Svalbard and Jan Mayen Islands
NCL	New Caledonia	SLB	Solomon Islands
NER	Niger	SLE	Sierra Leone
NFK	Norfolk Island	SLV	El Salvador
NGA	Nigeria	SMR	San Marino
NIC	Nicaragua	SOM	Somalia
NIU	Niue	SPM	St. Pierre and Miquelon
NLD	Netherlands	STP	Sao Tome and Principé
NOR	Norway	SUN	USSR
NPL	Nepal	SUR	Surinam
NRU	Nauru	SWE	Sweden
NTZ	Neutral Zone	SWZ	Swaziland
NZL	New Zealand	SYC	Seychelles
OMN	Oman	SYR	Syrian Arab Republic
OOO	escape code, see note 3	TCA	Turks and Caicos Islands
PAK	Pakistan	TCD	Chad
PAN	Panama	TGO	Togo
(PCI)	See FSM, MHL, MNP and PLW	THA	Thailand
PCN	Pitcairn	TKL	Tokelau
PER	Peru	TMP	East Timor
PHL	Philippines	TON	Tonga
PLW	Palau	TTO	Trinidad and Tobago
PNG	Papua New Guinea	TUN	Tunisia
POL	Poland	TUR	Turkey
PRI	Puerto Rico	TUV	Tuvalu

ISO • Alpha-3" code	Country	ISO • Alpha-3" code	Country
TWN	Taiwan, Province of China	VUT	Vanuatu
TZA	Tanzania, United Republic of	(WAK)	See UMI
UGA	Uganda	WLF	Wallis and Futuna Islands
UKR	Ukrainian SSR	WSM	Samoa
UMI	United States Minor Outlying Islands	XAA-XZZ	private use, see note 2
URY	Uruguay	YEM	Yemen
USA	United States	YMD	Yemen, Democratic
VAT	Vatican City State (Holy See)	YUG	Yugoslavia
VCT	Saint Vincent and the Grenadines	ZAF	South Africa
VEN	Venezuela	ZAR	Zaire
VGB	Virgin Islands (British)	ZMB	Zambia
VIR	Virgin Islands (U.S.)	ZWE	Zimbabwe
VNM	Viet Nam	ZZA-ZZZ	private use, see note 2

Notes (relevant only in the context of the EBU subtitle data exchange format):

1. Codes shown in brackets were formerly attributed to certain countries. The new codes should be used instead (e.g. UMI should always replace WAK).

2. Codes in the series AAA-AAZ, QMA-QZZ, XAA-XZZ, and ZZA-ZZZ should not be used in the CO field.

3. The escape code OOO should not be used in the CO field.

Bibliography

[1] *Information processing - Coded character sets for text communication*
Addendum to Part 2: Latin alphabetic and non-alphabetic graphic characters
ISO Standard 6937-2:1983/Add.1:1989.

[2] *Information processing - 8-bit single-byte coded graphic character sets*
Part 5: *Latin/Cyrillic alphabet*
Part 6: *Latin/Arabic alphabet*
Part 7: *Latin/Greek alphabet*
Part 8: *Latin/Hebrew alphabet*
ISO Standard 8859.

[3] *Specification of the systems of the MAC/packet family*
EBU document Tech. 3258, 1986.

[4] *EBU time-and-control codes for television tape-recordings (625-line television systems)*
EBU document 3097, 3rd edition, 1982.

[5] *Codes for the representation of names of countries*
ISO Standard 3166, 1988.

[6] *Main characteristics of a "level-2" teletext system for European broadcasting organizations using the fixed-format principle*
EBU document Tech. 3240, 1982.