EUROPEAN TELECOMMUNICATION STANDARD

ETS 300 628

September 1994

Source: ETSI TC-SMG

Reference: GSM 03.38

ICS: 33.060.30

Key words: Europeen digital cellular telecommunications system, Global System for Mobile communications (GSM)

European digital cellular telecommunications system (Phase 2); Alphabets and language-specific information (GSM 03.38)

ETSI

European Telecommunications Standards Institute

ETSI Secretariat

Postal address: F-06921 Sophia Antipolis CEDEX - FRANCE
Office address: 650 Route des Lucioles - Sophia Antipolis - Valbonne - FRANCE
X.400: c=fr, a=atlas, p=etsi, s=secretariat - **Internet:** secretariat@etsi.fr

Tel.: +33 92 94 42 00 - Fax: +33 93 65 47 16

Copyright Notification: No part may be reproduced except as authorized by written permission. The copyright and the foregoing restriction extend to reproduction in all media.

© European Telecommunications Standards Institute 1994. All rights reserved.

New presentation - see History box

Whilst every care has been taken in the preparation and publication of this document, errors in content, typographical or otherwise, may occur. If you have comments concerning its accuracy, please write to "ETSI Editing and Committee Support Dept." at the address shown on the title page.

Contents

Blank page

Foreword

This European Telecommunication Standard (ETS) has been produced by the Special Mobile Group (SMG) Technical Committee (TC) of the European Telecommunications Standards Institute (ETSI).

This ETS defines Alphabets and language-specific information for the European digital cellular telecommunications system (Phase 2). This ETS corresponds to GSM Technical Specification (GSM-TS) GSM 03.38 version 4.0.1.

The specification from which this ETS has been derived was originally based on CEPT documentation, hence the presentation of this ETS may not be entirely in accordance with the ETSI/PNE rules.

Reference is made within this ETS to GSM-TSs.

NOTE: TC-SMG has produced documents which give the technical specifications for the implementation of the European digital cellular telecommunications system. Historically, these documents have been identified as GSM Technical Specifications (GSM-TSs). These TSs may have subsequently become I-ETSs (Phase 1), or ETSs (Phase 2), whilst others may become ETSI Technical Reports (ETRs). GSM-TSs are, for editorial reasons, still referred to in current GSM ETSs.

Blank page

1 Scope

This technical specification defines the language-specific requirements for GSM. These are specific codepoints required by the SMS specifications which in turn are used not only for SMS (GSM 03.40, 03.41) but also for Unstructured Data (GSM 02.90) and may additionally be used for MMI (GSM 02.30).

The specifications for the DCE/DTE interface (GSM 07.05, 07.06) will also use the codes specified herein for the transfer of SMS data to an external terminal.

2 Normative references

This ETS incorporates by dated and 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 ETS only when incorporated in it by amendment or revision. For undated references, the latest edition of the publication referred to applies.

[1] GSM 01.04 (ETR 100): "European digital cellular telecommunication system (Phase 2); Definitions, abbreviations and acronyms".

[2] GSM 02.30 (ETS 300 511): "European digital cellular telecommunication system (Phase 2); Man-Machine Interface (MMI) of the Mobile Station (MS)".

[3] GSM 02.90 (ETS 300 549): "European digital cellular telecommunication system (Phase 2); Unstructured supplementary services operation - Stage 1".

[4] GSM 03.40 (ETS 300 536): "European digital cellular telecommunication system (Phase 2); Technical realization of the Short Message Service (SMS) Point to Point (PP)".

[5] GSM 03.41 (ETS 300 537): "European digital cellular telecommunication system (Phase 2); Technical realization of Short Message Service Cell Broadcast (SMSCB)".

[6] GSM 04.11 (ETS 300 559): "European digital cellular telecommunication system (Phase 2); Point-to-Point (PP) Short Message Service (SMS) support on mobile radio interface".

[7] GSM 04.12 (ETS 300 560): "European digital cellular telecommunication system (Phase 2); Short Message Service Cell Broadcast (SMSCB) support on the mobile radio interface".

[8] GSM 07.05 (ETS 300 585): "European digital cellular telecommunication system (Phase 2); Use of Data Terminal Equipment - Data Circuit terminating Equipment (DTE - DCE) interface for Short Message Service (SMS) and Cell Broadcast Service (CBS)".

[9] GSM 07.06 (ETS 300 586): "European digital cellular telecommunication system (Phase 2); Use of the V series Data Terminal Equipment - Data Circuit terminating Equipment (DTE - DCE) interface at the Mobile Station (MS) for Mobile Termination (MT) configuration".

3 Definitions and abbreviations

Definitions used in this specification are listed in GSM 01.04.

4 SMS Data Coding Scheme

The TP-Data-Coding-Scheme field, defined in GSM 03.40, indicates the data coding scheme of the TP-UD field, and may indicate a message class. The octet is used according to a coding group which is indicated in bits 7..4. The octet is then coded as follows:

Coding Group Bits 7..4	Use of bits 3..0
0000	Alphabet indication Unspecified message handling at the MS. Bits 3..0 indicate the alphabet as follows: 0000 Default alphabet 0001..1111 Reserved
0001..1110	Reserved coding groups
1111	Data coding/message class Bit 3 is reserved, set to 0. Bit 2 Message coding: 0 Default alphabet 1 8-bit data Bit 1 Bit 0 Message Class: 0 0 Class 0 0 1 Class 1 default meaning: ME-specific. 1 0 Class 2 SIM-specific message. 1 1 Class 3 default meaning: TE specific (see GSM TS 07.05)

Default alphabet indicates that the TP-UD is coded from the 7-bit alphabet given in subclause 6.2.1. When this alphabet is used, the characters of the message are packed in octets as shown in subclause 6.1.2.1.1, and the message can consist of up to 160 characters. The default alphabet shall be supported by all MSs and SCs offering the service.

8-bit data indicates that the TP-UD has user-defined coding, and the message can consist of up to 140 octets.

When a mobile terminated message is class 0 and the MS has the capability of displaying short messages, the MS shall display the message immediately and send an acknowledgement to the SC when the message has successfully reached the MS irrespective of whether there is memory available in the SIM or ME. The message shall not be automatically stored in the SIM or ME.

The ME may make provision through MMI for the user to selectively prevent the message from being displayed immediately.

If the ME is incapable of displaying short messages or if the immediate display of the message has been disabled through MMI then the ME shall treat the short message as though there was no message class, i.e it will ignore bits 0 and 1 in the TP-DCS and normal rules for memory capacity exceeded shall apply.

When a mobile terminated message is Class 1, the MS shall send an acknowledgement to the SC when the message has successfully reached the MS and can be stored. The MS shall normally store the message in the ME by default, if that is possible, but otherwise the message may be stored elsewhere, e.g. in the SIM. The user may be able to override the default meaning and select their own routing.

When a mobile terminated message is Class 2 (SIM-specific), a phase 2 (or later) MS shall ensure that the message has been transferred to the SMS data field in the SIM before sending an acknowledgement to the SC. The MS shall return a "protocol error, unspecified" error message (see GSM TS 04.11) if the short message cannot be stored in the SIM and there is other short message storage available at the MS. If all the short message storage at the MS is already in use, the MS shall return "memory capacity exceeded".

When a mobile terminated message is Class 3, the MS shall send an acknowledgement to the SC when the message has successfully reached the MS and can be stored, irrespectively of whether the MS supports an SMS interface to a TE, and without waiting for the message to be transferred to the TE. Thus the acknowledgement to the SC of a TE-specific message does not imply that the message has reached the TE. Class 3 messages shall normally be transferred to the TE when the TE requests "TE-specific" messages (see GSM TS 07.05). The user may be able to override the default meaning and select their own routing.

The message class codes may also be used for mobile originated messages, to provide an indication to the destination SME of how the message was handled at the MS.

The MS will not interpret reserved or unsupported values but shall store them as received. The SC may reject messages with a Data Coding Scheme containing a reserved value or one which is not supported.

5 Cell Broadcast Data Coding Scheme

The Cell Broadcast Data Coding Scheme indicates the intended handling of the message at the MS, the alphabet/coding, and the language (when applicable). The octet is used according to a coding group which is indicated in bits 7..4. The octet is then coded as follows:

Coding Group Bits 7..4	Use of bits 3..0
0000	Language using the default alphabet Unspecified handling at the MS Bits 3..0 indicate the language: 0000 German 0001 English 0010 Italian 0011 French 0100 Spanish 0101 Dutch 0110 Swedish 0111 Danish 1000 Portuguese 1001 Finnish 1010 Norwegian 1011 Greek 1100 Turkish 1101..1110 Reserved for European languages 1111 Language unspecified
0001..0100	Reserved for European Languages using the default alphabet, with unspecified handling at the MS.
0101..1110	Reserved coding groups

1111	Data coding / message handling Bit 3 is reserved, set to 0. Bit 2 Message coding: 0 Default alphabet 1 8-bit data Bit 1 Bit 0 Message Class: 0 0 No message class. 0 1 Class 1 user defined. 1 0 Class 2 user defined. 1 1 Class 3 default meaning: TE-specific (see GSM TS 07.05)

These codings may also be used for Unstructured SS Data and MMI/display purposes.

Messages using the default alphabet are coded with the 7-bit alphabet given in subclause 6.2.1. The message then consists of 93 user characters.

Messages using 8-bit data have user-defined coding, and will be 82 octets in length.

Class 1 and Class 2 messages may be routed by the ME to user-defined destinations, but the user may override any default meaning and select their own routing.

Class 3 messages will normally be selected for transfer to a TE, in cases where a ME supports an SMS/CBS interface to a TE, and the TE requests "TE-specific" cell broadcast messages (see GSM TS 07.05). The user may be able to override the default meaning and select their own routing.

6 Individual parameters

6.1 General principles

6.1.1 General notes

Except where otherwise indicated, the following shall apply to all alphabet tables:

1: The characters marked "1)" are not used but are displayed as a space.

2: The characters of this set, when displayed, should approximate to the appearence of the relevant characters specified in ISO 1073 and the relevant national standards.

3: Control characters:

Code Meaning

LF Line feed: Any characters following LF which are to be displayed shall be presented as the next line of the message, commencing with the first character position.

CR Carriage return: Any characters following CR which are to be displayed shall be presented as the current line of the message, commencing with the first character position.

SP Space character.

4: The display of characters within a message is achieved by taking each character in turn and placing it in the next available space from left to right and top to bottom.

6.1.2 Character packing

6.1.2.1 SMS Point-to-Point Packing

6.1.2.1.1 Packing of 7-bit characters

If a character number α is noted in the following way:

b7	b6	b5	b4	b3	b2	b1
αa	αb	αc	αd	αe	αf	αg

The packing of the 7-bits characters in octets is done by completing the octets with zeros on the left.

For examples, packing:

- one character in one octet:
  bits number:

7	6	5	4	3	2	1	0
0	1a	1b	1c	1d	1e	1f	1g

- two characters in two octets:
  bits number:

7	6	5	4	3	2	1	0
2g	1a	1b	1c	1d	1e	1f	1g
0	0	2a	2b	2c	2d	2e	2f

- three characters in three octets:
  bits number:

7	6	5	4	3	2	1	0
2g	1a	1b	1c	1d	1e	1f	1g
3f	3g	2a	2b	2c	2d	2e	2f
0	0	0	3a	3b	3c	3d	3e

- seven characters in seven octets:
  bits number:

7	6	5	4	3	2	1	0
2g	1a	1b	1c	1d	1e	1f	1g
3f	3g	2a	2b	2c	2d	2e	2f
4e	4f	4g	3a	3b	3c	3d	3e
5d	5e	5f	5g	4a	4b	4c	4d
6c	6d	6e	6f	6g	5a	5b	5c
7b	7c	7d	7e	7f	7g	6a	6b
0	0	0	0	0	0	0	7a

- eight characters in seven octets:
  bits number:

7	6	5	4	3	2	1	0
2g	1a	1b	1c	1d	1e	1f	1g
3f	3g	2a	2b	2c	2d	2e	2f
4e	4f	4g	3a	3b	3c	3d	3e
5d	5e	5f	5g	4a	4b	4c	4d
6c	6d	6e	6f	6g	5a	5b	5c
7b	7c	7d	7e	7f	7g	6a	6b
8a	8b	8c	8d	8e	8f	8g	7a

The bit number zero is always transmitted first.

Therefore, in 140 octets, it is possible to pack (140x8)/7=160 characters.

6.1.2.2 SMS Cell Broadcast Packing

6.1.2.2.1 Packing of 7-bit characters

If a character number α is noted in the following way:

b7	b6	b5	b4	b3	b2	b1
αa	αb	αc	αd	αe	αf	αg

the packing of the 7-bits characters in octets is done as follows

	Bit number							
	7	6	5	4	3	2	1	0
Octet number								
1	2g	1a	1b	1c	1d	1e	1f	1g
2	3f	3g	2a	2b	2c	2d	2e	2f
3	4e	4f	4g	3a	3b	3c	3d	3e
4	5d	5e	5f	5g	4a	4b	4c	4d
5	6c	6d	6e	6f	6g	5a	5b	5c
6	7b	7c	7d	7e	7f	7g	6a	6b
7	8a	8b	8c	8d	8e	8f	8g	7a
8	10g	9a	9b	9c	9d	9e	9f	9g
				.				
				.				
81	93d	93e	93f	93g	92a	92b	92c	92d
82	0	0	0	0	0	93a	93b	93c

The bit number zero is always transmitted first.

Therefore, in 82 octets, it is possible to pack (82x8)/7 = 93.7, that is 93 characters. The 5 remaining bits are set to zero as stated above.

6.2 Alphabet tables

This section provides tables for all the alphabets to be supported by SMS. The default alphabet is mandatory. Additional alphabets are optional. Irrespective of support of an individual alphabet, an MS shall have the ability to store a short message coded in any alphabet on the SIM.

6.2.1 Default alphabet

Bits per character: 7

SMS User Data Length meaning: Number of characters

CBS pad character: CR

Character table:

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	@	Δ	SP	0	¡	P	¿	p
0	0	0	1	1	£	1)	!	1	A	Q	a	q
0	0	1	0	2	$	Φ	"	2	B	R	b	r
0	0	1	1	3	¥	Γ	#	3	C	S	c	s
0	1	0	0	4	è	Λ	¤	4	D	T	d	t
0	1	0	1	5	é	Ω	%	5	E	U	e	u
0	1	1	0	6	ù	Π	&	6	F	V	f	v
0	1	1	1	7	ì	Ψ	'	7	G	W	g	w
1	0	0	0	8	ò	Σ	(	8	H	X	h	x
1	0	0	1	9	Ç	Θ	)	9	I	Y	i	y
1	0	1	0	10	LF	Ξ	*	:	J	Z	j	z
1	0	1	1	11	Ø	1)	+	;	K	Ä	k	ä
1	1	0	0	12	ø	Æ	,	<	L	Ö	l	ö
1	1	0	1	13	CR	æ	-	=	M	Ñ	m	ñ
1	1	1	0	14	Å	ß	.	>	N	Ü	n	ü
1	1	1	1	15	å	É	/	?	O	§	o	à

History

Document history	
September 1994	First Edition
October 1995	Converted into Adobe Acrobat Portable Document Format (PDF)