GSM TECHNICAL SPECIFICATION

GSM 03.38

March 1996

Version 5.1.0

Source: ETSI TC-SMG

Reference: TS/SMG-040338QR

ICS: 33.060.50

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

GSM

GLOBAL SYSTEM FOR
MOBILE COMMUNICATIONS

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

*

Contents

Blank page

Foreword

This Global System for Mobile communications Technical Specification (GTS) has been produced by the Special Mobile Group (SMG) Technical Committee (TC) of the European Telecommunications Standards Institute (ETSI).

This GTS defines the language-specific requirements for GSM within the digital cellular telecommunications system (Phase 2/Phase 2+).

This GTS is a TC-SMG approved GSM technical specification version 5, which contains GSM Phase 2+ enhancements/features to the version 4 GSM technical specification. The European Telecommunications Standard from which this Phase 2+ GTS has evolved is Phase 2 GSM ETS 300 628 (GSM 03.38 version 4.0.1).

GTS are produced by TC-SMG to enable the GSM Phase 2+ specifications to become publicly available, prior to submission for the formal ETSI standards approval procedure to become European Telecommunications Standards (ETS). This ensures the earliest possible access to GSM Phase 2+ specifications for all Manufacturers, Network operators and implementors of the Global System for Mobile communications.

The contents of this GTS are subject to continuing work within TC-SMG and may change following formal TC-SMG approval. Should TC-SMG modify the contents of this GTS it will then be republished by ETSI with an identifying change of release date and an increase in version number as follows:

Version 5.x.y

where:

- y the third digit is incremented when editorial only changes have been incorporated in the specification;

- x the second digit is incremented for all other types of changes, i.e. technical enhancements, corrections, updates, etc.

Reference is made within this TS to GSM-TSs (note).

NOTE: TC-SMG has produced documents which give the technical specifications for the implementation of the 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/ETSI Technical Reports (ETRs) (Phase 2). TC-SMG has also produced ETSI GSM TSs which give the technical specifications for the implementation of Phase 2+ enhancements of the digital cellular telecommunications system. These version 5.x.x GSM Technical Specifications may be referred to as GTSs.

Blank page

1 Scope

This Global System for Mobile communications Technical Specification (GTS) defines the language-specific requirements for GSM. These are specific codepoints required by the Short Message Service (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 Man Machine Interface (MMI) (GSM 02.30).

The specification for the Data Circuit terminating Equipment/Data Terminal Equipment (DCE/DTE) interface (GSM 07.05) will also use the codes specified herein for the transfer of SMS data to an external terminal.

2 Normative references

This GTS 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 GTS 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): "Digital cellular telecommunication system (Phase 2); Abbreviations and acronyms".

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

[3] GSM 03.90 (ETS 300 549): "Digital cellular telecommunication system (Phase 2); Unstructured supplementary services operation - Stage 1".

[4] GSM 03.40 (ETS 300 536): "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): "Digital cellular telecommunication system (Phase 2); Technical realization of Short Message Service Cell Broadcast (SMSCB)".

[6] GSM 04.11 (ETS 300 559): "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): "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): "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)".

[10] ISO/IEC10646: “Universal Multiple-Octet Coded Character Set (UCS)”; UCS2, 16 bit coding.

3 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:

<table>
<tr><td>Coding Group Bits 7..4</td><td>Use of bits 3..0</td></tr>
<tr><td>00xx</td><td>General Data Coding indication
Bits 5..0 indicate the following :

Bit 5, if set to 0, indicates the text is uncompressed
Bit 5, if set to 1, indicates the text is compressed using the GSM standard compression algorithm. (yet to be specified)

Bit 4, if set to 0, indicates that bits 1 to 0 are reserved and have no message class meaning
Bit 4, if set to 1, indicates that bits 1 to 0 have a message class meaning :

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)

Bits 3 and 2 indicate the alphabet being used, as follows :
Bit 3 Bit2 Alphabet:
0 0 Default alphabet
0 1 8 bit
1 0 UCS2 (16bit) [10]
1 1 Reserved

NOTE: The special case of bits 7..0 being 0000 0000 indicates the Default Alphabet as in Phase 2</td></tr>
<tr><td>1000..1011</td><td>Reserved coding groups</td></tr>
<tr><td>1100</td><td>Message Waiting Indication Group: Discard Message
Bits 3..0 are coded exactly the same as Group 1101, however with bits 7..4 set to 1100 the mobile may discard the contents of the message, and only present the indication to the user.</td></tr>
<tr><td>1101</td><td>Message Waiting Indication Group: Store Message
This Group allows an indication to be provided to the user about the status of types of message waiting on systems connected to the GSM PLMN. The mobile may present this indication as an icon on the screen, or other MMI indication. The mobile may take note of the Origination Address for messages in this group and group 1100. For each indication supported, the mobile may provide storage for the Origination Address which is to control the mobile indicator
Text included in the user data is coded in the Default Alphabet.
Where a message is received with bits 7..4 set to 1101, the mobile shall store the text of the SMS message in addition to setting the indication.
Bits 3 indicates Indication Sense:

Bit 3
0 Set Indication Inactive
1 Set Indication Active

Bit 2 is reserved, and set to 0

Bit 1 Bit 0 Indication Type:
0 0 Voicemail Message Waiting
0 1 Fax Message Waiting
1 0 Electronic Mail Message Waiting
1 1 Other Message Waiting*

* Mobile manufacturers may implement the “Other Message Waiting” indication as an additional indication without specifying the meaning. The meaning of this indication is intended to be standardised in the future, so Operators should not make use of this indication until the standard for this indication is finalised.</td></tr>
</table>

1110	Message Waiting Indication Group: Store Message The coding of bits 3..0 and functionality of this feature are the same as for the Message Waiting Indication Group above, (bits 7..4 set to 1101) with the exception that the text included in the user data is coded in the uncompressed UCS2 alphabet.
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.

UCS2 alphabet indicates that the TP-UD has a UCS2 [10] coded message, and the message can consist of up to 140 octets, i.e. up to 70 UCS2 characters.

When a message is compressed, the TP-UD consists of the default alphabet or UCS2 alphabet compressed message, and the compressed message itself can consist of up to 140 octets in total.

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 appearance 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: α

- 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

Octet number \ Bit number	7	6	5	4	3	2	1	0
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	
October 1995	Creation of Version 5.0.0 (Version 4.0.1 + AR 2)
December 1995	Publication of Version 5.0.0
February 1996	Creation of Version 5.1.0 (Version 5.0.0 + CR 3)
March 1996	Publication of Version 5.1.0

ISBN 2-7437-0581-7
Dépôt légal : Mars 1996