TS 100 900 V7.2.0 (1999-07)

Technical Specification

Digital cellular telecommunications system (Phase 2+); Alphabets and language-specific information (GSM 03.38 version 7.2.0 Release 1998)

GLOBAL SYSTEM FOR MOBILE COMMUNICATIONS

European Telecommunications Standards Institute

Reference
RTS/SMG-040338Q7 (8c003000.PDF)

Keywords
Digital cellular telecommunications system,
Global System for Mobile communications
(GSM)

ETSI Secretariat

Postal address
F-06921 Sophia Antipolis Cedex - FRANCE

Office address
650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16
Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Internet
secretariat@etsi.fr
http://www.etsi.fr
http://www.etsi.org

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available **free of charge** from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://www.etsi.org/ipr).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This ETSI Technical Specification (TS) has been produced by the Special Mobile Group (SMG) of the European Telecommunications Standards Institute (ETSI).

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

The contents of this TS is subject to continuing work within SMG and may change following formal SMG approval. Should SMG modify the contents of this TS, it will be re-released with an identifying change of release date and an increase in version number as follows:

Version 7.x.y

where:

- 7 indicates Release 1998 of GSM Phase 2+
- x the second digit is incremented for changes of substance, i.e. technical enhancements, corrections, updates, etc.
- y the third digit is incremented when editorial only changes have been incorporated in the specification.

1 Scope

This TS 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 [8]) will also use the codes specified herein for the transfer of SMS data to an external terminal.

2 Normative references

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies.
- A non-specific reference to an ETS shall also be taken to refer to later versions published as an EN with the same number.
- For this Release 1998 document, references to GSM documents are for Release 1998 versions (version 7.x.y).

[1] GSM 01.04: "Digital cellular telecommunication system (Phase 2+); Abbreviations and acronyms".

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

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

[4] GSM 03.40: "Digital cellular telecommunication system (Phase 2+); Technical realization of the Short Message Service (SMS) Point to Point (PP)".

[5] GSM 03.41: "Digital cellular telecommunication system (Phase 2+); Technical realization of Short Message Service Cell Broadcast (SMSCB)".

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

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

[8] GSM 07.05: "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.

[11] GSM 04.90: "Digital cellular telecommunication system (Phase 2+); Unstructured supplementary services operation - Stage 3".

[12] ISO 639 "Code for the representation of names of languages"

[13] GSM 03.42: "Digital cellular telecommunication system (Phase 2+); Compression algorithm for text messaging services".

[14] GSM 03.40: "Digital cellular telecommunication system (Phase 2+); Technical realization of the Short Message Service (SMS) Point to Point (PP)".

[15] "Wireless Datagram Protocol Specification", Wireless Application Protocol Forum Ltd.

3 Abbreviations

Abbreviations used in this TS 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. Any reserved codings shall be assumed to be the GSM default alphabet (the same as codepoint 00000000) by a receiving entity. 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
00xx	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. (see GSM TS 03.42) 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 [8]) Bits 3 and 2 indicate the alphabet being used, as follows : Bit 3 Bit2 Alphabet: 0 0 Default alphabet 0 1 8 bit data 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
0100..1011	Reserved coding groups
1100	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.

(continued)

Coding Group Bits 7..4	Use of bits 3..0
1101	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 standardized in the future, so Operators should not make use of this indication until the standard for this indication is finalized.
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 [8])

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. If the 7 bit default alphabet extension mechanism is used then the number of displayable characters will reduce by one for every instance where the 7 bit default alphabet extension table is used8-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. The General notes specified in subclause 6.1.1 override any contrary specification

in UCS2, so for example even in UCS2 a <CR> character will cause the MS to return to the beginning of the current line and overwrite any existing text with the characters which follow the <CR>.

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". $begin$(Secure SMS)$ This behaviour applies in all cases except for phase 2+ MS supporting SIM Application Toolkit when the Protocol Identifier (TP-PID) of the mobile terminated message is set to "SIM Data download" (see GSM 03.40 [14]).end(Secure SMS)$.

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 [8]). 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). Any reserved codings shall be assumed to be the GSM default alphabet (the same as codepoint 00001111) by a receiving entity. 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 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 Hungarian 1110 Polish 1111 Language unspecified
0001	0000 Default alphabet; message preceded by language indication. The first 3 characters of the message are a two-character representation of the language encoded according to ISO 639 [12], followed by a CR character. The CR character is then followed by 90 characters of text. A Pre-Phase 2+ MS will overwrite the start of the message up to the CR and present only the text. 0001 UCS2; message preceded by language indication The message starts with a two 7-bit default alphabet character representation of the language encoded according to ISO 639 [12]. This is padded to the octet boundary with two bits set to 0 and then followed by 40 characters of UCS2-encoded message. An MS not supporting UCS2 coding will present the two character language identifier followed by improperly interpreted user data. 0010..1111 Reserved for European languages
0010..	0000 Czech 0001 .. 1111 Reserved for European Languages using the default alphabet, with unspecified handling at the MS
0011	0000..1111 Reserved for European Languages using the default alphabet, with unspecified handling at the MS

(continued)

(concluded)

01xx	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 compressing algorithm. (see GSM TS 03.42) 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 [8]) Bits 3 and 2 indicate the alphabet being used, as follows: Bit 3 Bit 2 Alphabet: 0 0 Default alphabet 0 1 8 bit data 1 0 USC2 (16 bit) [10] 1 1 Reserved
1000..1101	Reserved coding groups
1110	Defined by the WAP Forum [15]
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 [8])

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

See GSM 04.90 [11] for specific coding values applicable to Unstructured SS Data for MS originated USSD messages and MS terminated USSD messages. USSD messages using the default alphabet are coded with the 7-bit alphabet given in subclause 6.2.1. The message can then consist of up to 182 user characters.

Cell Broadcast 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.

If the 7 bit default alphabet extension mechanism is used then the number of displayable characters will reduce by one for every instance where the 7 bit default alphabet extension table is usedCell Broadcast messages using 8-bit data have user-defined coding, and will be 82 octets in length.

UCS2 alphabet indicates that the message is coded in UCS2 [10]. The General notes specified in subclause 6.1.1 override any contrary specification in UCS2, so for example even in UCS2 a <CR> character will cause the MS to return to the beginning of the current line and overwrite any existing text with the characters which follow the <CR>. Messages encoded in UCS2 consist of 41 characters.

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 07.05 [8]). 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:

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.

6.1.2.3 USSD packing

6.1.2.3.1 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-bit characters in octets is done by completing the octets with zeros on the left.

For example, 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

- six characters in six 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
0	0	0	0	0	0	6a	6b

- 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	1	1	0	1	7a

The bit number zero is always transmitted first.

- 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

- nine characters in eight 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
0	9a	9b	9c	9d	9e	9f	9g

- fifteen characters in fourteen 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
10g     9a  9b  9c  9d  9e  9f  9g
11f 11g     10a     10b     10c     10d     10e     10f
12e 12f 12g     11a     11b     11c     11d     11e
13d     13e     13f 13g     12a     12b     12c     12d
14c 14d     14e     14f 14g     13a     13b     13c
15b     15c     15d     15e     15f 15g     14a     14b
0   0   0   1   1   0   1   15a
```

- sixteen characters in fourteen 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
10g     9a  9b  9c  9d  9e  9f  9g
11f 11g     10a     10b     10c     10d     10e     10f
12e 12f 12g     11a     11b     11c     11d     11e
13d     13e     13f 13g     12a     12b     12c     12d
14c 14d     14e     14f 14g     13a     13b     13c
15b     15c     15d     15e     15f 15g     14a     14b
16a 16b     16c     16d     16e     16f 16g     15a
```

The bit number zero is always transmitted first.

Therefore, in 160 octets, is it possible to pack (160*8)/7 = 182.8, that is 182 characters. The remaining 6 bits are set to zero as stated above.

Packing of 7 bit characters in USSD strings is done in the same way as for SMS (subclause 7.1.2.1).The character stream is bit padded to octet boundary with binary zeroes as shown above.

If the total number of characters to be sent equals (8n-1) where n=1,2,3 etc. then there are 7 spare bits at the end of the message. To avoid the situation where the receiving entity confuses 7 binary zero pad bits as the @ character, the carriage return or <CR> character (defined in subclause 7.1.1) shall be used for padding in this situation, just as for Cell Broadcast.

If <CR> is intended to be the last character and the message (including the wanted <CR>) ends on an octet boundary, then another <CR> must be added together with a padding bit 0. The receiving entity will perform the carriage return function twice, but this will not result in misoperation as the definition of <CR> in subclause 7.1.1 is identical to the definition of <CR><CR>.

The receiving entity shall remove the final <CR> character where the message ends on an octet boundary with <CR> as the last character.

Under certain circumstances, a Pre Phase 2 + MS will perform the carriage return function after displaying the last USSD character received.

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, a 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/USSD 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	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

1) This code is an escape to an extension of the 7 bit default alphabet table. A receiving entity which does not understand the meaning of this escape mechanism shall display it as a space character.

6.2.1.1 GSM 7bit default alphabet extension 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					|			
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							2)	
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	3)							
1	0	1	1	11		1)						
1	1	0	0	12				[				
1	1	0	1	13				~				
1	1	1	0	14				]				
1	1	1	1	15			\					

In the event that an MS receives a code where a symbol is not represented in the above table then the MS shall display the character shown in the main default 7 bit alphabet table in section 6.2.1

1) This code value is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined.

2) This code represents the EURO currency symbol. The code value is that used for the character ‘e’. Therefore a receiving entity which is incapable of displaying the EURO currency symbol will display the character ‘e’ instead.

3) This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile which does not understand the 7 bit default alphabet table extension mechanism will treat this character as Line Feed

6.2.2 8 bit data

8 bit data is user defined

SMS User Data Length meaning: Number of octets

Padding: CR in the case of an 8 bit character set

Otherwise - user defined

Character table: User Specific

6.2.3 UCS2

Bits per character: 16

SMS User Data Length meaning: Number of octets

CBS/USSD pad character: CR

Character table: ISO/IEC10646 [10]

Annex A (Informative): Document change history

SMG#	TDoc	SPEC	VERS	NEW_VERS	CR	REV	PHASE	CAT	WORKITEM	SUBJECT
s25	096/98	03.38	5.6.0	6.0.0	A015		R97		SIM toolkit security	Class 2 SIM Data download message handling
s26	291/98	03.38	6.0.0	7.0.0	A016		R98		TEI	7 bit default alphabet extensions
s28	99-061	03.38	7.0.0	7.1.0	A017		R98		TEI	changes for CBS 8 bit data and CBS compression
s29	99-482	03.38	7.1.0	7.2.0	A018		R98	C	MExE R98	Data Coding Scheme for WAP over USSD and CB

History

Document history	
July 1999	