Standardization Supporting Cultural Diversity: Character Repertoires, Ordering and Assignment to the 12-key Telephone Keypad for European Languages and Languages Used in Europe

Martin Böcker[1], Karl Ivar Larsson[2] and Bruno von Niman[3] (ETSI STF 300)

[1] *BenQ Mobile, Germany*
[2] *LWP Consulting, Sweden*
[3] *vonniman consulting, STF 300 Leader, Sweden*
bruno@vonniman.com

Abstract

ETSI, the European Telecommunications Standards Institute, has published a standard (ES 202 130) in 2003 that specifies the character repertoires and assignment of characters on the 12-key telephone keypad for a range of European languages. The standard for letters, digits and special characters (such as the Euro symbol and punctuation marks) covered the official languages of the EU and EFTA members, Russia, as well as countries with applicant status for the EU at that time. This paper describes the further development of the standard to cover other major languages spoken in Europe including official languages, minority languages and immigrants' languages.

Key words: Usability, user interfaces, standards, 12-key keypad, ICT

1. Introduction

Telecommunications devices currently represent one of the largest consumer product segments. As telecommunications devices and services converge with technologies such as information processing, broadcast services and the internet, while at the same time becoming mobile and ubiquitous, the usability of these devices and services becomes a critical factor in service uptake. One of the most challenging aspects of mobile-device usability is text entry using the standard 12-key telephone keypad.

At present, finding the characters necessary to enter a name in the terminal's phone book, searching for a name, writing an SMS (text) message or logging on to a mobile internet portal cannot always be performed easily, because manufacturers differ in terms of which European characters their devices support, how they are ordered in lists and how the specific characters are mapped onto the keys of the keypad. Character-set implementation varies sometimes even between devices and applications from one and the same manufacturer. Standardizing the way characters are mapped onto keypads gives users easier access to different communication de-

vices and services, allowing simple, correct and efficient text input, search and retrieval. It also broadens market opportunities for manufacturers and suppliers and reduces their development costs.

The original reason for assigning letters to the rotary dial pad and later to the numeric telephone keys was to provide alphabetic 'aliases' for digits, as mnemonics in dialling. The need to use a telephone keypad for entering text or data was not envisaged. Nobody in the pioneer days of telephony anticipated the concept of a 'phone books' stored inside the telephone, or a service like SMS, the very successful service for transmitting short text messages as an alternative to voice communication.

The only standards previously available (e.g. ETSI ETS 300 640 or ITU-T Recommendation E.161 (02/01)), addressing the assignment of characters to the 12-key telephone keypad, were limited to the assignment of the basic 26 Latin letters (a to z). Language-specific letters (e.g. ü, é, å, ä, ö) as well as other characters (e.g. '€' or '@') were not addressed. The lack of a standard on these issues has led to diverse and inconsistent solutions for European languages, obviously creating accessibility barriers to basic communication access in *e*Europe.

Europe has around 230 indigenous languages – worldwide there are close to 7000. The largest number of languages presently supported by a specific ICT device or service is approaching 50. Cultural and linguistic diversity is one of the key strengths of Europe. However, in ICT, it raises issues that need to be considered and solved in order not to limit access to services, their availability and usability, on the basic as well as more advanced levels.

The first version of ETSI ES 202 130 has been developed to solve the problem for some of the most important European languages by defining character repertoires, sorting orders and the assignment of letters to the 12-key telephony keypad for these languages. A new version of ETSI ES 202 130 will extend this work to cover other major languages spoken in Europe including official languages, minority languages and immigrants' languages. All of this work was aligned with the European Commission's initiative ***eEurope***, a programme for accelerated uptake and inclusive deployment of new, important, consumer-oriented technologies (http://europa.eu.int/information_society/eeurope).

2. Scope of ETSI ES 202 130 on its presently published form

The current version of ETSI ES 202 130 specifies the minimum repertoire and assignment of graphic (letter, digit and special) characters to standard 12-key telephone keypads on ICT devices with telephony functionality. It applies to public or private, fixed or mobile network terminals, without an alphanumeric keyboard but providing a 12-key keypad in hardware form (e.g. as push button keys) or software form (e.g. as soft keys on a visual display). It also applies to network-based services accessed through such terminal devices. It complements ETS 300 640 by additionally including European language-specific letters (Latin, Greek and Cyrillic scripts) and other common characters (e.g. '€' and punctuation marks). It specifies solutions for both language-independent and language-specific keypad assignments, mapped to the 12-key telephone keypad, also providing common and language-specific information on character repertoires and ordering.

The standard is fully applicable to the official languages of the European Union (EU) member countries as of 2005 and those of countries with candidate status (Romania, Bulgaria and Turkey) and, additionally, to the official languages of the EFTA (the European Free Trade

Association) countries and Russian. The languages fully covered by the first version of ETSI ES 202 130 are therefore: Bulgarian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Icelandic, Irish, Italian, Latvian, Lithuanian, Luxemburgish, Maltese, Norwegian, Polish, Portuguese, Romanian, Russian, Slovak, Slovenian, Spanish, Swedish, and Turkish. In anticipation of future expansions, the language-independent repertoires and keypad assignments specified also include letters needed in some of the remaining European official languages.

ETSI ES 202 130 does not cover any implementation related issues, e.g. specifics of predictive text input or user interface design.

3. User requirements

Users of the standard are those implementing it, for example interaction designers and other developers of ICT devices and services, designing user interfaces deploying text input and output, applied to 12-key keypad arrays provided in hardware form (e.g. as push button keys) or software form (e.g. as soft keys on a visual display) and telecommunication-network based services accessed through such terminal devices.

End users addressed are the consumers (end users) of the ICT devices and services mentioned above, ranging from first time to experienced advanced users, who can produce tactile stimuli in the form of a key press and perceive written text. The end users' main goal is to efficiently use ICT devices and services under circumstances intended by these. The implementation of ES 202 130 enables users to reapply knowledge and previous experience between different ICT devices and services using a 12-key standard keypad array and a display. Control of common functions such as entering of characters and retrieval of text in a certain order will be simplified. Well-established services which rely on alpha mnemonics (e.g . '800 DOCTOR' rather than '800 362867' are not negatively influenced as the standard only complements ETS 300 640).

For certain end users with special needs, ES 202 130 is particularly helpful due to consistent implementations (same character always found in the same position, regardless of the terminal manufacturer). For certain disabilities, e.g. in the case of temporary or permanent difficulties caused by cognitive problems or the lack of necessary level of proficiency in the respective language and other communication impairments such as: visual impairments, the inability to produce distinctive tactile stimuli or difficulties in handling, distinguishing and understanding textual information, the standard is not expected to have any impact.

Uniformity in the basic interactive elements increases the transfer of learning between devices and services and improves the overall usability of the entire interactive environment. Such transference becomes even more important in a world of ubiquitous devices and services.
Guiding principles during the development of the ordering and assignments of the alphanumeric characters have been:

1) Consistent and harmonised across different devices and services
2) Easy to learn and remember
3) As natural as possible, matching previously acquired knowledge
4) Redundancy (multiple solutions possible to reach desired input)

4 Methodology

4.1 Initial survey

As start of the work to developing the standard, an informal survey of the key assignments in a number of mobile phone models was carried out on several major manufacturers' handsets. The survey was based mainly on specifications and user manuals downloaded from the internet but also on 'hands-on' investigation.

4.2 Principles applying to ES 202 130

In order to arrive at a consistent and easy-to-implement presentation of the requirements for character repertoires, ordering rules and character assignment to the 12-key keypad, the principles listed in Table 1 were applied throughout the production of the standard. Some of these are elaborated in the following.

4.3 Characters needed

Approximately 240 Latin-repertoire letters are needed to cover the major European languages. With Greek and Cyrillic letters added, the number increases to well over 350. This can be compared to the 75 Latin-repertoire letters (mix of capital and small) supported by the present GSM 03.38 7-bit scheme generally implemented in today's mobile phones and networks (85 letters all-in-all when the Greek capital letters of that scheme are included). It was found necessary to include in the language-specific repertoires more letters than are contained in the "core" of those languages, called "Type A" letters. This is because in all languages there is a user need to input also foreign-origin words, some of them needing "foreign" letters. Further, in all countries there exist user preferences in spelling of some names with "foreign" letters, and possibly also a need to represent names – personal and/or geographical – correctly in recognised minority languages. The repertoire tables therefore also include "Type B" letters (see Figure 1).

4.4 Character ordering

Ordering of characters is a highly complex problem, and has been the subject of very large amounts of work in several standardisation bodies, both national and international. Earlier ETSI and ISO/IEC standards specify principles based on a "multi-level" approach for the ordering of strings of characters. However, it was found necessary to adopt a simplified "single-level" method for this standard, considering the limited capabilities of telephone devices as compared to computer systems. As regards letters, the two language-independent repertoire tables specify a deterministic ordering. For the language-specific repertoire tables, however, some additional criteria were applied because of established practices in telecommunications, e.g. for printed telephone directories.

In all European languages, the letters A-Z are considered part of the alphabet even if, in many of them, some of the letters are not used in any indigenous-origin words. Also some languages have special-shape letters, like the Icelandic Þ and the German ß. Additionally, all languages use special variants of letters A-Z with diacritical marks, like the acute accent and the cedilla (e.g. É and Ç). For ordering, most languages consider such variants equivalent to the basic letter. In some languages, however, a few of them are considered letters of their own, and ordered differently. For instance, the letter Ö is ordered in Swedish as the last letter of the alphabet. As far as possible, national conventions were followed for the language-specific repertoire tables. This may possibly cause "non-deterministic" ordering in specific cases. Although unsatisfactory in principle, it was concluded that this could be accepted for the relevant applications.

Table 1: Principles employed in character repertoires, sorting order and keypad assignment

Principle 1: Presentation of character repertoires and sorting orders

- Combine repertoire and ordering information in one table
- Provide language-independent tables per script (Latin, Cyrillic, Greek)
- Cover in language-independent tables languages not covered in language-specific part by designing the language-independent tables to be "future-proof" (e.g. Ukrainian, Serbian and Croatian)

Principle 2: Character description

- Describe letters in terms of standardized identifiers: (a) Letter: Representation of the letter, (b) GSM 03.38 7-bit coding, (c) ISO/IEC 6937 coding, (d) ISO/IEC 10646 (Unicode) identifier, and (e) ISO/IEC 10646 (Unicode) name
- Order characters according to established standards, e.g. the Latin and Cyrillic language-independent repertoires are ordered according to ENV 13710

Principle 3: Language-independent repertoires

- Latin: covers all Latin-based letters covered by the scope of the document
- Cyrillic: Repertoire according to ISO/IEC 8859-5:1998
- Greek-script repertoire is identical with the Greek language-specific repertoire
- Provide minimum Latin subset ("A – Z") to be used with the Cyrillic and Greek repertoires

Principle 4: Language-specific repertoires

- List essential alphabet of a particular language and letters typically used in that language (from various recognised sources). Usage type: A classification of each letter according to the following principles: (A) Letters essential to the language, and (B) Letters commonly used in writing the language, but not essential for it

Principles 5: Repertoire of digits and special characters

- Only one (European) language-independent table of digits and special characters is provided
- The need for language-specific tables is to be discussed
- The digits and special characters are ordered (at present) according to ISO/IEC 14651 resp. CEN ENV 13710

Principle 6: Information contained in the keypad assignment tables

- Key: the key of the 12-key keypad the respective letters are assigned to, Letter: Representation of the letter, ISO/IEC 10646 (Unicode) identifier, and ISO/IEC 10646 (Unicode) name

Principle 7: Latin-script assignment principles

- If a character is assigned to a key of the 12-key keypad, it shall be assigned to the key specified in the respective table
- Letters with diacritical marks are assigned to the same key of the 12-key keypad as their respective basic letters (if existent), i.e. "ä" is assigned to key "2" because "a" is assigned to "2" according to ITU-T E.161
- A character may be additionally assigned to other keys
- Complete language-independent and language-specific tables may be implemented in any combination

Principle 8: Greek-language and Cyrillic-script assignment principles

- The Greek-language repertoire and the Cyrillic-script repertoire shall be assigned together with the minimum Latin-script repertoire
- Additional characters not covered by the present document may be assigned to a key
- Only tables for the assignment of small letters are specified, capital letters shall be assigned in the same way as the respective small letter

Principle 9: Assignment order for Latin-script letters

- Letters assigned to that particular key according to ITU-T E.161 (e.g. "abc" to key "2")
- The digit for the respective key according to ITU-T E.161
- Type A letters according to the tables in Section 6 (e.g. "ä" on key "2" for German)
- Type B letters according to the tables in Section 6 (e.g. "à" on key "2" for German) (e.g. the resulting assignment for key "2" for German is "abc2ää")

Principle 10: Language-independent Latin-script assignment

- Letters are assigned to the above-mentioned principles and ordered according to ISO/IEC 14651 resp. CEN ENV 13710

Principle 11: Assignment order for Greek-script and Cyrillic-script letters

- Letters assigned to that particular key in alphabetic order (e.g. "абвг" to key "2")
- The digit for the respective key according to ITU-T E.161
- Latin letters assigned to that particular key according to ITU-T E.161 (e.g. abc to key "2") For example, the resulting assignment for key "2" for Russian is "абвг2abc")

Principle 12: Character ordering for Greek-language and the Cyrillic-script tables

- The characters of the Greek-language and the Cyrillic-script tables are ordered according to ISO/IEC 14651 resp. CEN ENV 13710

Letter	GSM 03.38 7-bit coding	ISO/IEC 6937 coding	ISO/IEC 10646 identifier	ISO/IEC 10646 name	Usage type	Notes
a	6/01	06/01	U+0061	LATIN SMALL LETTER A	A	•
A	4/01	04/01	U+0041	LATIN CAPITAL LETTER A	A	•
á	—	12/02 06/01	U+00E1	LATIN SMALL LETTER A WITH ACUTE	A	•
Á	—	12/02 04/01	U+00C1	LATIN CAPITAL LETTER A WITH ACUTE	A	•
ä	7/11	12/08 06/01	U+00E4	LATIN SMALL LETTER A WITH DIAERESIS	B	
Ä	5/11	12/08 04/01	U+00C4	LATIN CAPITAL LETTER A WITH DIAERESIS	B	
b	6/02	06/02	U+0062	LATIN SMALL LETTER B	A	
B	4/02	04/02	U+0042	LATIN CAPITAL LETTER B	A	
c	6/03	06/03	U+0063	LATIN SMALL LETTER C	A	‡
C	4/03	04/03	U+0043	LATIN CAPITAL LETTER C	A	‡
č	—	12/15 06/03	U+010D	LATIN SMALL LETTER C WITH CARON	A	
Č	—	12/15 04/03	U+010C	LATIN CAPITAL LETTER C WITH CARON	A	
d	6/04	06/04	U+0064	LATIN SMALL LETTER D	A	•
D	4/04	04/04	U+0044	LATIN CAPITAL LETTER D	A	•
ď	—	12/15 06/04	U+010F	LATIN SMALL LETTER D WITH CARON	A	•
Ď	—	12/15 04/04	U+010E	LATIN CAPITAL LETTER D WITH CARON	A	•
e	6/05	06/05	U+0065	LATIN SMALL LETTER E	A	•

Figure 1: Extract of the table specifying character repertoire and sorting order for Czech

Key	Letter	ISO/IEC 10646 identifier	ISO/IEC 10646 name
2	a	U+0061	LATIN SMALL LETTER A
	b	U+0062	LATIN SMALL LETTER B
	c	U+0063	LATIN SMALL LETTER C
	2	U+0032	DIGIT TWO
	á	U+00E1	LATIN SMALL LETTER A WITH ACUTE
	č	U+010D	LATIN SMALL LETTER C WITH CARON
	ä	U+00E4	LATIN SMALL LETTER A WITH DIAERESIS
3	d	U+0064	LATIN SMALL LETTER D
	e	U+0065	LATIN SMALL LETTER E
	f	U+0066	LATIN SMALL LETTER F
	3	U+0033	DIGIT THREE
	ď	U+010F	LATIN SMALL LETTER D WITH CARON
	é	U+00E9	LATIN SMALL LETTER E WITH ACUTE
	ě	U+011B	LATIN SMALL LETTER E WITH CARON
4	g	U+0067	LATIN SMALL LETTER G
	h	U+0068	LATIN SMALL LETTER H
	i	U+0069	LATIN SMALL LETTER I
	4	U+0034	DIGIT FOUR
	í	U+00ED	LATIN SMALL LETTER I WITH ACUTE

Figure 2: Extract of the table specifying character assignment to 12-key keypad for Czech

4.5 Keypad input sequences

In today's keypad-input implementations – foremost in mobile phones – the digits are generally placed as the last character in the key-press sequence, following not only the standardized letter assignments (ABC on key 2, DEF on key 3 etc.) but also all special letter variants assigned to the keys. The same principle was considered for ES 202 130. However, the special needs of visually-impaired users make the principle questionable. It was, therefore, decided to place, instead, the digits immediately following the presently standardized letter assignments; i.e. as the fourth key-press on all keys except 7 and 9 (PQRS and WXYZ) where it is the fifth (see Table 1 and Figure 2).

4.6 Digits and special characters

As the ordering and keypad assignment of digits and special characters turned out to be somewhat controversial, they were treated following a different set of rules. ES 202 130 defines a set of special characters that must be supported. In addition, other characters may also be supported. The order of appearance in the respective table is only a recommendation, valid for a language-independent implementation, and alternative orders of appearance of special characters are allowed. Furthermore, language-specific orders of appearance are also allowed. The full set of special characters must be accessible via one single entry point. It is recommended that this entry point is the "1" key. In addition, a device may use different other keys to access different sets of special characters and/or digits. In this case, Rule 1 and Rule 6 must still be followed. Thereby, the possibility to implement language-specific keypad assignments of special characters and digits is made possible.

5 Update and extension of ES 202 130

ES 202 130 was met with positive responses from industry. An extension of its language coverage is therefore highly desirable, in particular in view of the strong emphasis on multilingualism by the European Union. In January 2006, ETSI decided to begin work on such an extension. This will take the form of either a revision of ES 202 130, or of a complementary standard. The Terms of Reference for the decided-on work specifies the extension as containing "major minority languages, some official European languages and…non-European languages used by a considerable number of ICT users in Europe". The interpretation of this is not obvious, since different delimitations may be concluded from the wording.

Geographically, "Europe" is traditionally delimited in the south and the east by the Bosphorus, the Caucasian mountain range, the Ural Mountains and the Ural River. The North Atlantic islands (but generally not Greenland) are also included, as well as those Mediterranean islands not in the proximity of the African continent. This definition is however unsatisfactory as a language basis for the Terms of Reference, since it contains only a small part of the nation of Turkey, also part of Kazakhstan, but not Cyprus, and further leaving the Trans-Caucasian states somewhat undefined. A more suitable definition of "Europe" should be the one of the Council of Europe (CoE), containing all "traditionally European" states, and also clarifying that Turkey and Cyprus are part of Europe, but that Kazakhstan is not. Also CoE concluded after a thorough investigation, considering historical and cultural as well as other factors, that the Trans-Caucasian states Georgia, Armenia and Azerbaijan shall be considered European, and therefore eligible for entry in the Council (and all three are nowadays members).

An adoption of this definition for the update/extension work does not necessarily imply that all of the CoE member states' official/majority languages will be covered in the decided-on extension/standard. In particular, the Armenian and Georgian unique script systems will have to be considered, and will be studied in the initial phases of the standard extension work. A complication of the CoE definition is that it includes all of the Russian state, which is obviously European as well as Asiatic. This, however, relates only to the selection of minority languages to be covered in the ETSI work, and not to the definition as such. Another complication is the overseas territorities of some of the European states. This will need study, which will be performed in the initial stage of the work. As regards European-origin minority languages, special consideration will be taken of languages recognised in ratifications of the CoE charter ETS 148, "European Charter for Regional or Minority Languages". This charter has so

far been ratified by about half of the CoE member states, and signed – although not yet ratified – by several others. The ES 202 130 update/complement could therefore cover:

- official/majority languages of European countries not covered in the current version of the standard (e.g. Croatian);
- recognised European-country minority languages not already covered by the majority language (e.g. Sorbian in Germany)
- other important but not officially recognised minority languages (e.g. Friulian in Italy);
- large immigrant languages (e.g. Arabic);
- other important immigrant non-European languages (e.g. Vietnamese, which poses special character complications).

6 Summary

The implementation of ES 202 130 in its current version allows users to enter text into modern ICT devices in a number of major European languages, in a way that is logically consistent and that renders the learning of new keypad assignments superfluous when moving from the devices of one manufacturer to those of another. The usability of future ICT devices will be further increased by a revision/complement of ES 202 130 to expand the number of languages covered by the standard.

References

All ETSI references are available free of charge at www.etsi.org

CEN ENV 13710 (2000): *"European Ordering Rules - Ordering of characters from the Latin, Greek and Cyrillic scripts".*

ETSI ES 202 130 *Human Factors; User Interfaces; Character repertoires, ordering and assignment to the 12-key telephone keypad (European languages)*

ETSI ETS 300 640 *Human Factors (HF); Assignment of alphabetic letters to digits on standard telephone keypad arrays*

ETSI TS 100 900 *Digital cellular telecommunications system (Phase 2+); Alphabets and language-specific information (same as GSM 03.38 version 7.2.0, Release 1998)*

ISO 8859-7: *"Information processing - 8-bit single-byte coded graphic character sets - Part 7: Latin/Greek alphabet".*

ISO/IEC 10646-1 (2000): *"Information technology - Universal Multiple-Octet Coded Character Set (UCS); Part 1: Architecture and Basic Multilingual Plane".*

ISO/IEC 6937 (2001): *"Information technology - Coded graphic character set for text communication - Latin alphabet".*

ITU-T Recommendation E.161 (02/01) *Arrangement of digits, letters and symbols on telephones and other devices that can be used for gaining access to a telephone network.*