

DISPLAYABLE CHARACTER SETS FOR BROADCAST TELETEXT

Tech. 3232-E

Second edition - June 1982

CONTENTS

| | | |
|-------------|--|----|
| 1 | Introduction..... | 3 |
| 2. | EBU inquiry..... | 4 |
| 3. | Analysis of the replies from EBU members..... | 4 |
| 3.1 | Minimum number of characters required..... | 4 |
| 3.2 | Two or more languages on one page..... | 6 |
| 4. | Establishment of multilingual character repertoires..... | 8 |
| 4.1 | Constraints that have to be taken into account..... | 8 |
| 4.2 | Significance of the repertoires..... | 8 |
| 4.3 | Character repertoires for languages written in Latin-based-alphabets..... | 9 |
| 4.4 | The combination of non-Latin-based alphabets together with a limited selection of Latin-based characters - | 11 |
| 5. | Conclusion..... | 14 |
| | Bibliographical references..... | 15 |
| Appendix 1- | Example of the forms used in the EBU inquiry..... | 17 |
| Appendix 2- | The complete repertoire of Latin-based characters required by EBU Member-organisations..... | 18 |
| Appendix 3- | Generation of the characters of the ASV-CODAR/I set by means of a 5 x 9 dot matrix. | 20 |
| Appendix 4- | Coding table for a combined repertoire of ASV-CODAR/I and certain Latin-based characters. | 21 |
| Appendix 5- | Provisional proposed combined repertoire for the Arabic alphabet. | 22 |
| Appendix 6- | Provisional proposed combined repertoire for the Cyrillic alphabet..... | 23 |
| Appendix 7- | Provisional proposed combined repertoire for the Greek alphabet..... | 24 |
| Appendix 8- | Provisional proposed combined repertoire for the Hebrew alphabet..... | 25 |
| Appendix 9- | Generation of the characters specific to Hebrew and the figures of Indian origin, in 5 x 9 dot matrix. | 26 |

Following the preparation of the first edition of this document, in 1980, information has been obtained on certain additional aspects of the requirements of EBU Member-organisations, and this has been taken into account in preparing this second edition. The complete Latin-based repertoire has been revised, and proposals are given for four "combined repertoires", each consisting of a combination of the characters required for the display of the seven languages most widely used in Western Europe with those required for the display of texts in one of the following non-Latin-based alphabets: Arabic, Cyrillic, Greek and Hebrew. As a number of amendments to the original text have been made also, this text should be considered to replace that of the first edition.

DISPLAYABLE CHARACTER SETS FOR BROADCAST TELETEXT

1 Introduction

The topic of data transmission systems in broadcasting has been attracting more and more attention during the past few years, and the Bureau of the EBU Technical Committee decided, at its meeting in October 1976, that an Ad-hoc group should be formed to study the characteristics of such systems and, in the case of teletext, the possibilities of defining a standard European alphabet and of drafting the specifications of a universal receiving terminal.

The character sets employed by the teletext services that have already been introduced were derived from the set standardised by the ISO and CCITT for professional data-processing, by the addition and replacement of a few special characters. However, experience has shown that these sets are in general inadequate and inefficient for applications in domestic teletext broadcasting. On the one hand, public acceptance of this new medium of communication requires a character set at least as extensive as that customary on typewriters intended for the languages concerned, and on the other hand the limited number of characters that can be displayed on a single "page" makes it necessary to be able to use as many abbreviations and symbols as possible.

Because no reliable information was available on the precise desires of the EBU Member-organisations in this field, the Ad-hoc Group, which has since become known as Sub-group V2, therefore undertook a survey of their minimum requirements regarding character repertoires for use in domestic data broadcasting services. The survey was organised in such a way that information could be obtained on both the total set of characters from which it was desired that any page of text could be composed, and the sets of characters specific to each language, so that the requirements for different combinations of languages could be derived.

This document consists of an analysis of the replies received in response to the survey, together with proposals for various sets of displayable characters, known as *repertoires*, complying with the requirements expressed by EBU Member-organisations. It is intended that this information should provide the basis for international standardisation in the field of teletext broadcasting.

2. EBU inquiry

At the end of 1978, the EBU undertook a survey of the requirements of its Members in the field of character sets for broadcast teletext. In order to facilitate the analysis of the replies, respondents were asked to indicate, for each language likely to be required for teletext, the minimum character set necessary, together with what other languages were likely to be required on the same page. The replies were to be entered on a form containing two tables, each of which was subdivided into eight different types of character. One table already contained the proposed basic character set (consisting of all the characters common to the character sets initially specified for U.K. teletext and Antiope), and the other already contained examples of the additional characters required for one of four geographically-related groups of languages. Examples of the forms used are reproduced in *Appendix 1*. The respondents were asked to delete all redundant characters in both tables, and to insert any additional characters required in the second table, so that the contents of both tables taken together would represent the minimum set required for the language under consideration.

3. Analysis of the replies from EBU members

A very large response was obtained; nearly all EBU Active Members returned completed forms, and several of them also provided explanations of the reasons for which their choice was made. As in some cases several different organisations provided information regarding the same language, it is also possible to obtain some idea of the degree of flexibility in the requirements.

3.1 Minimum number of characters required

Table 1 shows the total minimum number of characters required for each language by each broadcasting organisation, and how many are included in the "basic" and "additional" categories. It can be seen that each of the Northern European languages and Turkish require a total of no more than approximately 96 characters (except in the cases of Dutch and Norwegian, which are discussed below). On the other hand, the Romance languages (French, Italian, Spanish and Portuguese) all require substantially more than 96 characters, generally between 110 and 127; the exceptionally large requirements of Italian (RAI) are due to the inclusion of characters which are taken from other alphabets or are restricted to particular applications. It should be noted that the information about Spanish given in *Table 1* (but not elsewhere) corresponds to the combined requirements for Basque, Catalan, Galician and Spanish. Finally, so far as languages written in non-Latin-based alphabets are concerned, whereas Arabic requires a total of less than 96 characters, Greek requires 127 characters if all accented characters, and the accents themselves, are to be included.

Table 1.- Summary of the essential requirements for character sets

| Reference | Language | Organisation | Other languages requested on same page | No. of characters | | |
|--|-------------------------------|-------------------------|--|-------------------|------------|------------|
| | | | | Basic | Additional | Total |
| a) Languages written in Latin-based alphabets | | | | | | |
| 1 | Basque | FR3 | 9 | 89 | 12 | 101 |
| 2 | Breton | FR3 | 9 | 89 | 11 | 100 |
| 3 | Catalan | FR3 | 9 | 89 | 19 | 108 |
| 4 | Croatian Serbian | JRT | 18,25,26,33 | 87 | 14 | 101 |
| 5 | Danish | DR | | 83 | 13 | 96 |
| 6 | Dutch | BRT FR3 NOS | 7,9,10 | 89 | 24 | 113 |
| | | | 9 | 89 | 16 | 105 |
| | | | 4,7,10,12,22,30* | 89 | 17 | 106 |
| 7 | English | BBC/UKIBA CTV RTE | - | 90 | 6 | 96 |
| | | | - | 90 | 8 | 98 |
| | | | 11 | 90 | 6 | 96 |
| 8 | Finnish Swedish | YLE | * | 83 | 9 | 92 |
| 9 | French | A2 | 6,7,10,12,17,19,22,30 | 90 | 30 | 120 |
| | | FR3 | 1,2,3,6,10,12,16 | 89 | 26 | 115 |
| | | RTBF | 6,10,23 | 90 | 37 | 127 |
| | | RTL | - | 90 | 18 | 108 |
| | | RTM | - | 83 | 6 | 89 |
| | | RTT | - | 90 | 22 | 112 |
| | | SRF | - | 90 | 21 | 111 |
| | | SSR | 10,12 | 90 | 19 | 109 |
| | | TFL | 6,7,10,12,17,19,30 | 89 | 25 | 114 |
| TMC | - | 90 | 37 | 127 | | |
| 10 | German | ARD | 4,6,7,9,12,17,19,22 | 84 | 11 | 95 |
| | | FR3 | 9 | 89 | 14 | 103 |
| | | ORF | - | 89 | 8 | 97 |
| | | RTL | - | 90 | 3 | 93 |
| | | SRG | 9,12 | 90 | 4 | 94 |
| | | ZDF | 4,6,7,9,12,17,19,22 | 85 | 9 | 94 |
| 11 | Irish | RTE | 7 | 74 | 17 | 91 |
| 12 | Italian | FR3 | 9 | 89 | 22 | 111 |
| | | RAI | 7 9 7,9,10,19* | 90 | 20 | 110 |
| | | TSI | 9,10 | 90 | 20 | 110 |
| 13 | Lapp | YLE, | 8 | 83 | 15 | 98 |
| 14 | Norwegian | NRK | 13* | 83 | 40 | 123 |
| 15 | Icelandic | | | | | |
| 16 | Occitan | FR3 | 9 | 89 | 21 | 110 |
| 17 | Portuguese | RTP | - | 90 | 31 | 121 |
| 18 | Slovenian | JRT | 25,26,33,34 | 87 | 10 | 97 |
| 19 | Spanish | RTVE | * | 88 | 31 | 119 |
| 20 | Basque Catalan Galician | | | | | |

* It is desired also to be able to compose texts in the other main languages of Western Europe that are written in Latin-based alphabets.

According to the reply from the RAI regarding Italian, for example, this would require a total of 168 characters.

Table 1.- Summary of the essential requirements for character sets (cont'd)

| Reference | Language | Organisation | Other languages requested on same page | No. of characters | | |
|--|-------------|--------------|--|-------------------|------------|-------|
| | | | | Basic | Additional | Total |
| b) Languages written in Latin-based alphabets | | | | | | |
| 21 | Swedish | SR | * | 79 | 10 | 89 |
| 22 | Turkish | TRT | - | 79 | 12 | 91 |
| 23 | Wallon | RTBF | 9 | 86 | 7 | 110 |
| 24 | Welsh | BBC UKIBA | 7 | 89 88 | 8 | 96 |
| 25 | Albanian | - | | | | |
| 26 | Hungarian | - | | | | |
| b) Languages not written in Latin-based alphabets | | | | | | |
| 30 | Arabic | RTA | | 20 | 47 | 67 |
| | | RTM | | 30 | 64 | 94 |
| | | RTT | | 20 | 71 | 91 |
| 31 | Greek | CyBC | | 57 | 34 | 91 |
| | | ERT | | 35 | 92 | 127 |
| 32 | Hebrew | IBA | 30* | 31 | 30 | 61 |
| 33 | Macedonian | JRT | 18,25,26,34 | 52 | 49 | 101 |
| 34 | Serbo-Croat | JRT | 4,18,25,26,33 | 48 | 52 | 100 |

* It is desired also to be able to compose texts in the other main languages of Western Europe that are written in latin-based alphabets.

According to the reply from the F~I regarding Italian, for example, this would require a total of 168 characters.

3.2 Two or more languages on one page

In several cases, however, it is considered to be necessary to make provision for use of two or more languages on the same page, and therefore the minimum number of characters required if a common repertoire is employed in such cases must also be examined.

3.2.1 Languages written in Latin-based alphabets

First of all, there are the countries in which there are two or more languages of roughly equal importance. In the case of Switzerland, for example, the reply from SSR indicates that it should be possible to write French, German and Italian on the same page. On the basis of the use of a common character repertoire for all three languages, a total of at least 121 characters would be needed. Furthermore, both the character sets required for Belgium by the BRT and RTBF include provision for writing German as well as the two principal national languages, Dutch and French. In Ireland, too, both Irish and English may be required on any page.

Next, there are those countries in which one or more additional languages are used by a minority of their inhabitants, so that while it must be possible to display texts in the majority language on pages composed in the minority language, the converse is not necessary. For example, in the United Kingdom, Welsh is in this position with respect to English. Similarly, according to the reply from YLE, texts in Finnish and Swedish (which make use of the same character set) may have to be displayed on the same page as those in the Lapp alphabet (98 characters). Although no extra characters are needed in this case, the possibility of adding ü and Ü for use in foreign names is desired (100 characters). A further requirement is taken into account in the reply for Norwegian (NRK), which would require only some 96 characters when those needed for Lapp are included, as requested, whereas 123 are needed to ensure the possibility of writing foreign words, such as names, in their original spelling.

Finally, the broadcasting organisations in some other countries also require the possibility of writing texts in several foreign languages on the same broadcast page. In the case of Dutch (NOS), French (A2, FR3 and TF1) and German (ARD and ZDF), these include the principal languages of Western Europe, and in the case of Swedish, SR expressed a strong desire to have 17 additional characters, and to retain the existing practice of spelling foreign words, such as names, in their original form, which would require a further 62, mainly accented, characters. In these four countries, the use of teletext pages and sub-titling in Serbo-Croat, Turkish etc. for the benefit of immigrants is also foreseen. A similar requirement for the possibility of writing the principal languages of Western Europe and the minority languages of Slovenian and Ladin, together with a need for various special characters intended mainly for mathematical applications, is given in the case of Italian by the RAI.

3.2.2. Languages not written in Latin-based alphabets

The question of accommodating the requirements for the display of texts in more than one language on the same page also arises in the case of the languages written in alphabets that are not based on the Latin alphabet: Arabic. Cyrillic. Greek and Hebrew. Although this was not specifically required by all the broadcasting organisations concerned, the possibility of defining a *combined repertoire* containing not only all the characters required for each non-Latin-based alphabet but also a selection of characters from various Latin-based alphabets sufficient to permit texts in the languages most widely used in Western Europe to be displayed on the same page too has been considered, with the aim of facilitating international and educational applications. As will become apparent in § 4.3, the total number of characters required for such combined repertoires must exceed 128 if an adequate selection of Latin-alphabet-based characters is included; the various aspects of this question are discussed in detail and solutions in the case of the above-mentioned alphabets are proposed in § 4.4.

4. Establishment of multilingual character repertoires

4.1 Constraints that have to be taken into account

The problem that has to be solved is that of defining and arranging one or more multilingual character repertoires which permit the greatest flexibility in use, while complying with the various technical and linguistic constraints. Before examining the options available, it is first necessary to consider the possible constraints:

a) Technical constraints

The principal constraints that must be taken into account are due to the need to transmit the characters by "words" in binary code. For this reason, repertoires not exceeding, for example, 96, 128 or 224 characters are preferable. Maximum compatibility with character repertoires contained in existing standards such as CCITT Recommendation 8.61 should be maintained also, and attention paid to the resolution available for displaying each character on the television screen.

b) Linguistic constraints

The various requirements for intermingled use of two or more languages on the same page must be satisfied. Furthermore, the possibility of introducing additional language options of this kind should be taken into account, specifically with regard to the likelihood of cross-frontier television reception ("spillover"). The risk of ambiguity when exploiting the possibility of multiple use of individual characters such as "-", as both hyphen and minus operator, must be avoided.

To some extent, the relative importance of these constraints must be weighted arbitrarily. Furthermore, it may be possible to give advantages to one character repertoire only at the cost of introducing disadvantages to another; the implications of all constraints must therefore be examined from this aspect.

4.2 Significance of the repertoires

It should be appreciated that the repertoires indicate only the functions of the various characters, and they do not define the style in which they should be displayed. However, in the case of the "combined repertoires", advantage has been taken of the fact that certain characters are common to both of the alphabets included in them to minimise the total number of characters to be generated, and therefore care should be taken in the design of character generators to ensure that such characters blend harmoniously with both the alphabets concerned. Manufacturers of character generators are otherwise free to adopt whichever 'font' they consider to be most appropriate. In the case of certain non-Latin-based alphabets, however, an indication has been given of how the characters might be generated on the basis of a 5 x 9 dot matrix, in order to provide guidance on what is considered by the broadcasting organisations concerned to be the most satisfactory solution in the case of low-resolution displays.

It should be noted that, in some cases, it may be possible to increase the acceptability of texts transmitted in accordance with such repertoires by systematically displaying certain characters in an alternative style. For example, when the figures 0 to 9 are associated with text in Arabic characters they could be replaced, at the discretion of the individual user, by the corresponding characters of Indian origin used in certain

Arabian countries. Some cases where such possibilities exist for enhancing the appearance of displayed text without requiring the inclusion of additional characters in the repertoire used for transmission are discussed briefly in the corresponding sections of this document.

4.3 Character repertoires for languages written in Latin-based-alphabets

The fundamental constraint is the limitation on the number of characters in the repertoire, and the options available for different values of this parameter are discussed below. In particular, the possibility of devising multilingual repertoires satisfying the minimum requirements of the most widely-used languages that employ Latin-based alphabets is examined.

a) 96 characters

According to the replies received, the essential requirements of 9 languages used by EBU Active Members serving 194 million people could each be satisfied with a character set not exceeding 96 characters. On the other hand, the essential requirements of 5 languages used by EBU Active Members serving 161 million people could not be satisfied with sets of only 96 characters. Furthermore, the only cases in which it is possible to satisfy the requirements for a combined character set allowing texts in two or more languages to be written on the same page, without exceeding 96 characters, are Finnish/Swedish + Lapp and English + Welsh. The adoption of a maximum set of 96 characters would thus not meet the operational requirements of many EBU Active Members.

b) 128 characters

According to the replies received, the essential requirements of each of the 24 languages used by EBU Active Members could be satisfied by a character set not exceeding 128 characters. Furthermore, it would be possible to satisfy virtually all the essential requirements of all the seven most widely used languages, covering 337 million people in 11 Western European countries, by adding 35 characters to the 90 contained in the basic character set given in the questionnaire; the remaining undesignated characters could then be employed independently in each language for functions such as accented capital letters, which are not considered to be essential for international communications. This full 125-character repertoire, which covers the requirements for Dutch, English, French, German, Italian, Spanish and Turkish, and is referred to hereafter as the "common core", is shown in *Table 2*.

Alternatively, Portuguese (10 million potential users) could be taken as the seventh language instead of Turkish (136 million potential users). Although Turkey, unlike Portugal, has no common frontier with any country using a Latin-based "alphabet, so that no consideration of overlap zones is necessary, the total number of immigrant Turkish workers and their families in other European countries is probably similar to, if not greater than, the number of Portuguese expatriates in them.

In the case of the Nordic countries, each reply (except that from Danmarks Radio) expressed the desire to have the possibility of writing names etc. correctly in other European languages. The possibility of writing texts in other Nordic languages was not specifically required, but it seems to be reasonable to give this priority, particularly as the number of potential users for the individual Nordic languages is relatively small, by devising a combined Nordic character repertoire incorporating the requirements of Danish, Finnish, Icelandic, Norwegian and Swedish, which have a total

potential of some 26 million users. Such a set would require eight characters additional to those in the above-mentioned common-core, and it could thus be accommodated within the limit of 128 characters by use of the three unassigned characters and by replacement of those assigned to the arrows and the symbols for 'greater than' and 'lesser than', which are not required in any Nordic character set.

Table 2. - Proposed 125-character "common core"

(German, English, French, Italian, Turkish, Spanish, Dutch)
approximately 337 million potential users in Western Europe

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|------------------------------|---|---|---|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|---|---|---|---|---|---|---|---|---|---|---|--|--|
| 1. Lower-case letters | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z | ç | ğ | ı | ñ | ş | ß | | |
| 2. Upper-case letters | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | Ç | İ | Ñ | Ş | Û | | | |
| 3. Punctuation/abbreviations | ! | " | # | \$ | % | ' | (|) | * | £ | , | - | . | / | : | ; | < | > | ? | - | & | ¿ | € | | | | | | | | | | | |
| 4. Accented letters | á | à | â | ö | é | è | ê | ë | í | ì | î | ï | ó | ò | ô | õ | ú | ù | û | ü | ij | | | | | | | | | | | | | |
| 5. Numerals | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | | | | | | | | | | | | | | | | | | | | | | | | |
| 6. Mathematical symbols | + | ÷ | = | ° | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7. Pictograms | ← | → | ↑ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

c) 222 characters; the complete Latin-based repertoire

As has already been mentioned, the use of separate 128-character sets would allow the essential requirements of each of the principal Western European languages using Latin-based alphabets to be satisfied in all cases, and it would also enable a common-core set sufficient to meet the needs of all the seven languages most widely used in Western Europe to be established. It would not be possible, however, to provide simultaneously both the common-core and the characters from a non-Latin-based alphabet required for writing a language such as Arabic or Greek, without exceeding 128 characters. Furthermore, the full character set required for Italian by the RAI (193 characters, certain of which are not strictly essential, although they would be needed to enable teletext to be used in many additional applications) could not be accommodated within that total. No doubt experience with the use of teletext will lead other broadcasting organisations to demand a more extensive character set, incorporating the additional characters needed for specialised applications, but not involving the extra complications of the use of dynamically redefinable character sets (DRCS). As the next step after 128 in a binary system is 256, and some 34 codes must be reserved for control functions, etc., the definition of multi-lingual repertoires not exceeding 222 characters should be considered.

A particularly useful repertoire would be one containing every character proposed in the requirements for the individual languages written in Latin-based alphabets, as full universal compatibility between all teletext transmissions in Western Europe could be obtained only by its use. If the 25 additional specialised characters (mathematical symbols) required by the RAI are disregarded, the total number of different characters required by

EBU Member-organisations amounts to 221 characters. This "complete Latin-based repertoire", which is given in *Appendix 2*, consists of a definitive list of all those characters that EBU Member-organisations have indicated to be required for teletext broadcasting using Latin-based alphabets in the following languages: Basque, Breton, Catalan, Croatian, Danish, Dutch, English, Finnish, French, Galician, German, Icelandic, Irish, Italian, Lapp, Norwegian, Occitan, Portuguese, Serbian, Slovenian, Spanish, Swedish, Turkish, Walloon and Welsh. It is presented in the same format as § 3.2.2 of CCITT Rec. S.100, with the same reference numbers where applicable. The characters of the EBU repertoire that do not appear in Rec. S.61 may be distinguished by the absence of reference numbers.

4.4 The combination of non-Latin-based alphabets together with a limited selection of Latin-based characters -

4.4.1 Principles for establishing combined repertoires

In order to enable proposals to be prepared for a repertoire consisting of a combination of a selection of characters from Latin-based alphabets with those required for displaying texts in a language written in a non-Latin-based alphabet, it is necessary first to establish the order of priority for incorporating characters into the repertoire. The most appropriate order is considered to be that in which the characters having the least general application in the area where the repertoire is to be used are given the lowest priority. Thus in preparing the proposals given in *Appendices 5. 6. 7 and 8.* for combined repertoires for the Arabic, Cyrillic, Greek and Hebrew alphabets respectively, the following decreasing order of priority has been adopted in determining which characters should be included in each repertoire:

- a) The set of 125 characters known as the "EBU common-core set". as defined in § 4.3b and illustrated in *Table 2*.
- b) The additional characters required by EBU Member-organisations for displaying texts in the languages using the non-Latin-based alphabet concerned, in conjunction with the common-core.
- c) The additional characters needed to enable texts to be displayed in the other languages required for use on the same page, in accordance with the corresponding reply to the EBU questionnaire.
- d) The additional characters needed to enable texts to be displayed in the official language(s) of neighbouring countries to that in which the EBU Member-organisation(s) requiring the language concerned is (are) located.
- e) The additional characters needed to enable texts to be displayed in other European languages using Latin-based alphabets, in decreasing order of the number of persons in the European Broadcasting Area having that language as their maternal language, so far as is known, without exceeding a total of 222 characters.

Although the arrangements for establishing coding tables for the transmission of these characters are not considered here, it is important to note that the possibility of using decoders intended for the display of transmissions in one language to display intelligible text when receiving transmissions in other languages - and thus the potential market for such decoders - can be maximised by ensuring that as few characters as possible are assigned different transmission codes in the different repertoires, and that these characters are of relatively low priority. For this reason, the proposed repertoires are sub-divided in accordance with the order of priority defined above.

4.4.2. Proposals for certain combined repertoires

a) Arabic

The composition and transmission of texts in Arabic by teletext involves a number of problems that do not arise in the case of those in Latin-based alphabets. In particular, Arabic is written from right to left, and it is customary to vary the form of certain letters according to their position in the word [1]. So far as the first aspect is concerned, it may reasonably be assumed that only the text composition equipment should be adapted accordingly, as the speed at which the text is displayed is high enough to prevent any inconvenience to readers if the conventional left-to-right transmission and display sequence is retained. With regard to the character repertoire, it would be appropriate to adopt, for *transmission* purposes, a set of approximately 64 characters specific to Arabic having the maximum practicable compatibility with the existing standard known as "ASV-CODAR/1" [2], which unfortunately does not include all the characters required by the EBU Members concerned, together with characters from the Latin-based alphabets selected according to the order of priority defined in § 4.4.1 above, up to a total of 222 characters. Alternatively, a repertoire identical with that proposed, in § 4.4.2 d) below could be obtained by replacing the 27 Latin-based characters of lowest priority by those specific to Hebrew.

The 64 Arabic characters given in ASV-CODAR/1 include only the "initial or medial" forms of the letters. If it is desired to reproduce the traditional variations in the form of the letters while displaying Arabic texts, as is required by certain EBU Members, the additional characters needed must be provided in the character-generator, and additional logic circuits must be provided in the decoder, such that the appropriate form is chosen, by means of analysis of the position that the letter occupies in the word, in response to the code corresponding to the form of the letter shown in the transmission repertoire. If it is sufficient to provide for two forms of each letter, one corresponding to the initial or medial form and the other to the "final or isolated" form, the logic circuit in the teletext decoder can be very simple, because the rows of text are transmitted starting at the left-hand end. In this case, the form in which each Arabic letter should be displayed depends on the nature of the character transmitted immediately before it: if it is a "space" generating character, the final or isolated form should appear, otherwise the initial or medial form should be displayed. The characters of the ASV-CODAR/1 set are shown in a 5 x 9 dot-matrix form in *Appendix 3*. It should be noted that an 8-bit coding table for a combination of ASV-CODAR/1 and the 96 Latin-based characters of ISO 646 has already been defined [2]; it is reproduced in *Appendix 4*. Studies with the aim of defining the set of characters in the transmission repertoire specific to Arabic are being undertaken in collaboration with the Arab States Broadcasting Union.

Although the requirements indicated by the EBU Member-organisations concerned do not include the versions of the figures 0 to 9 of Indian origin used in certain Arabian countries, it is desirable to consider how provision may be made for displaying them. Fortunately the convention for writing multi-digit numbers with these characters is the same as that with the figures 0 to 9 included in the common-core set, and thus it is sufficient to make provision in the teletext decoder to generate whichever set of figures is preferred by the user, in response to the same ten transmitted codes. The choice of the set of figures to be displayed should, of course, be made by the individual viewer, for example-by means of a switch on the decoder. This alternative set of 10 figures, together with the 64 Arabic characters of ASV-CODAR/1 and the 125 Latin-based characters of the EBU common-core, is included in the provisional combined repertoire

shown in *Appendix 5*. An indication of how the ten figures could be generated using a 5 x 9 dot matrix is given in *Appendix 9*.

b) Cyrillic

According to a contribution to the CCIR from Yugoslavia [3], 99 characters are required for the display of the Serbo-Croat language in the Cyrillic alphabet, and 101 characters for the display of the Macedonian language in that alphabet. However, 59 characters of the combined set of 105 characters required for these two languages are already included in the EBU common-core set of 125 characters, one is a mathematical symbol and another is a pictogram, leaving a total of 44 characters specific to Cyrillic. In other words, if the requirements of Serbo-Croat and Macedonian in the Cyrillic alphabet are combined with the 125 characters of the EBU common-core, a set having a total of 171 different characters results. There is; thus no difficulty in adding the extra characters needed for the display of texts in Albanian, Hungarian and Slovenian, as required by [3], without exceeding the maximum of 222 characters. As no information on the corresponding character sets required for displaying texts in these three languages is given in [3], it has been derived from a study undertaken by an expert from Rennes University [4]. Furthermore, in response to requirements notified after the main analysis of the replies had been completed, a total of 20 additional characters have been included in order to enable texts to be composed in the minority languages of Ruthenian, Romanian and Slovakian. As can be seen, however, it has not been possible to include these additional characters, nor the capital letters with two acute accents used in Hungarian, in the list of 221 Latin-based characters given in *Appendix 2*. Finally, although it was not specifically required to be able to display texts in the Russian language by means of this repertoire, only 6 additional Cyrillic characters would be needed for this purpose, and they have therefore been included in the proposed repertoire, which is reproduced as *Appendix 6*.

c) Greek

The requirements for displaying texts in the Greek language can be satisfied by a combination of 90 characters specific to Greek with the 125 characters of the EBU common-core Latin-based set. This total of 215 characters allows another 7 characters to be added without exceeding the maximum of 222. According to the principles defined in § 4.4.1 above, priority should be given to Latin-based alphabets used for the official languages of neighbouring countries, and thus it is proposed that the lower-case characters needed for displays in Serbian, Croatian and Slovenian should be added. The corresponding proposed repertoire is given in *Appendix 7*.

Note: Legislation enacted in Greece early in 1982 abolished the use of the existing three accents and introduced the use of a single new accent. The foregoing text and proposal will have to be modified to take account of this development.

d) Hebrew

This language is written from right-to-left; it may, however, be assumed that this characteristic has no influence on the display requirements. According to the information supplied by IBA, 27 characters specific to Hebrew are required in combination with 31 characters already included in the EBU common-core set and two fractions. The combined total of $27+125+2 = 154$ characters required for Hebrew and the seven most widely-used Western European languages having Latin-based alphabets allows a total of 68 characters to be added without exceeding the maximum of 222.

In accordance with the requirements indicated by IBA, provision for the display of texts in Arabic should also be made, and therefore it is proposed that the Arabic characters defined in § 4.4.2a should be included in the repertoire. These characters, of which there are approximately 64, do not, however, include the versions of the figures 0 to 9 of Indian origin used in certain Arabian countries. and therefore provision for displaying them, if desired, should be made. Means for doing this are proposed in § 4.4.2a. The choice of the set of figures to be displayed should, of course, be made by the individual user, for example by means of a switch on the decoder. The corresponding 10 characters are included in the proposed combined repertoire for Hebrew, which otherwise consists of a total of approximately 218 characters and is given in *Appendix 8*. Indications of how the characters specific to Hebrew may be generated by means of a 5 x 9 dot matrix are given in *Appendix 9*. (See also the remark in § 4.4.2a regarding the display of the various forms of Arabic letters.)

5. Conclusion

Information on the minimum contents of the character repertoires required for teletext broadcasting texts in each of 25 Western European languages that are written in Latin-based alphabets, and for that in certain languages written in Arabic, Cyrillic, Greek: and Hebrew alphabets, has been obtained from the EBU Members concerned. The principles on which combined repertoires permitting the minimum requirements for displaying texts in more than one of these languages to be satisfied can be prepared have been developed, and various proposals for such combined repertoires: have been put forward, taking into account the relevant literary and technical constraints. In particular, a repertoire of 221 characters enabling the requirements for composing texts in the Latin-based alphabets used for the following languages to be satisfied is given: Basque, Breton, Catalan, Croatian, Danish, Dutch, English, Finnish, French, Galician, German, Icelandic, Irish, Italian, Lapp, Norwegian, -Occitan, Portuguese, Serbian, Slovenian, Spanish, Swedish, Turkish, Walloon and Welsh. This may be considered to be a definitive statement of the EBU Members' requirements for these languages. Proposals are also given for four combined repertoires, each of which consists of a combination of the Latin-based characters required for the display of the seven languages most widely used in Western Europe with those required for the display of texts in languages using one of the following non-Latin-based alphabets: Arabic, Cyrillic, Greek and Hebrew. It is intended that these proposals should be taken as the basis for discussions leading to the establishment of definitive repertoires for broadcast teletext in the languages concerned.

Bibliographical references

[1] Mitchell, T.F. and Barbel~, D.: *Introduction to Arabic*
BBC Publications, London" 1972.

[2] ASV-CODAR: *Arabe standard voyellé - codage arabe* (Standard Arabic with vowels - Arabic coding).

Published by the Institut: d'études et de recherches pour l'arabisation,
Rabat (Morocco).

[3] *Displayable character sets for broadcast teletext.*
CCIR Document 11/136 (Study Period 1978-1982).

[4] *Problèmes linguistiques du télétexte : Etude des langues européennes utilisant l'alphabet latin - Rapport no 1* (Linguistic problems of teletext: Study of European languages using the Latin alphabet -Report No.1.
Published by the CCETT, Rennes, June 1977.