

TITLE OF INVENTION

METHOD FOR THE DISPLAY OF TELETEXT WITH LANGUAGE IDENTIFICATION

5

CROSS-REFERENCE TO RELATED APPLICATIONS

10 The application presented here is based upon and
claims priority from prior Singapore Patent Application No.
200801249, titled *Method of Language Coding Identification
and Data Format*. The entire disclosure of which is herein
incorporated by reference.

15

BACKGROUND OF THE INVENTION

Field of the Invention

20 The present invention relates to the field of Teletext
broadcast, which remains an important part of Television
services in many countries. It deals especially with the
determination and selection of different national languages
to display Teletext correctly.

25

Description of the Related Art

30 Teletext is a television information retrieval service
and it offers a range of text-based information, typically
national, international and sporting news and TV schedules.
Subtitles information is also transmitted in Teletext signal.
Teletext information is broadcast in the vertical blanking
interval between image frames in a broadcast television
signal. It is closely linked to the PAL broadcast systems, and
35 is nearly ubiquitous across Europe as well as other regions,
with most major broadcasters providing a Teletext service.

Naturally, as Teletext is text-based services, in a particular country the native language needs to be localized to provide the Teletext services in the local language. In a multi-lingual society or continent like Europe, it is essential to have more than one language supported. The Teletext Standard ETS 300 706 titled as "Enhanced Teletext Specification" has defined the fixed alphanumeric and graphic character sets referenced by ETS which to be used by Teletext in practice. By this standard, Teletext Level 1.5 is required to support most European languages (English, Czech/Slovak, Estonian, French, German, Italian, Lettish/Lithuanian, Polish, Portuguese/Spanish, Rumanian, Serbian/Croatian/Slovenian, Swedish/Finnish, Turkish), plus Arabic and Hebrew. It is worth noting that Level 1.5 is still the most popular Teletext standard in analogue and DVB transmissions in Europe today.

ETS 300 706 has defined character designation requirements to address the multi-language designation. In general, the standard has defined a primary Latin character set (G0 set) and a supplementary character set (G2), each of 96 entries, for each alphabet. The G0 set comprises the characters of the alphabet used to write the contents of the pages. Some characters in the G0 primary set may be substituted by a group of characters that are specific to a particular national language. These characters are grouped into a set called Latin National Option Sub-sets according to the national languages they are specific to. The national languages are: Czech/Slovak, English, Estonian, French, German, Italian, Lettish/Lithuanian, Polish, Portuguese/Spanish, Rumanian, Serbian/Croatian/Slovenian, Swedish/Finnish and Turkish. G2 supplementary set of characters are defined when the local language requirements require more than 96 alphanumeric characters to provide a basic service. In this

case, some characters from the G0 primary set will be incorporated.

5 The default G0 primary and G2 supplementary
character sets to be used by all pages in a magazine are
defined by the Default G0 and G2 Character Set Designation
and National Option Selection value (triplet 1, bits 8 to 14)
of a packet M/29/0 or M/29/4, or, with higher priority for a
particular page, by the same bits in a packet X/28/0 Format
10 1 or X/28/4.

 Teletext Level 1 transmissions are restricted to using
the G0 set and the Latin national option sub-sets. The
national option sub-sets that are to be selected to display
15 Teletext page are decided by the control bits C12, C13 and
C14 contained in the thirteenth (13th) eight-bit byte in the
page header packet. For example, C12 C13 and C14 which
have the value 0 0 0 means that this Teletext page is to be
written in English, while 0 1 1 means the language is Italian.
20 Typically, a Teletext decoder will look into the control bits
C12 C13 and C14 for its value to decide which national
language option sub-set to be selected for correctly decoding
and displaying this page.

25 The standard also specifies that at Teletext Level 1
and 1.5 the national option sub-set in use on the page is
defined by control bits C12, C13 and C14 alone, therefore in
theory, this will result in an ambiguous reference to the
national languages. For example, when C12 C13 and C14
30 have the value of 1 1 0, it could designate Czech/Slovak or
Turkish; while when C12 C13 and C14 have the value of 1 1
1, it either designates null or Rumanian.

 In order to address this problem, Level 2.5 and 3.5
35 have defined explicitly by bits in a packet X/28/0 Format 1
or X/28/4. However, as for Teletext Level 1 and 1.5 or when

X/28/0 Format 1 or X/28/4 triplet 1 bits 8 to 10 are absent, the national language designation control bits can cause some confusion when they deal with certain national languages.

5

In view of the above, there is a need to overcome the ambiguity by detecting and identifying the national language clearly and successfully so that a Teletext decoder could refer to the correct national language option sub-set for correct display of Teletext.

10

SUMMARY OF THE INVENTION

This invention proposes using a method to identify the correct language that is used in Teletext pages, and to select automatically and correctly this national language option subset fonts to be used for displaying it, when control bits C12, C13 and C14 designate confusing results or causing ambiguity.

20

The method comprises the following steps. The TV receiver or Teletext decoder receives the TV signal with Teletext service. If the decoder finds Teletext services, it will apply the single-byte language coding identification mechanism presented in the cross-reference Singapore Patent Application No. 200801249, to select which G0 primary set to be used. Based on the identification result, Latin G0 character set, or Cyrillic option 1, option 2 or option 3 will be selected as the main G0 character set. Once the G0 character set is selected, the control bits C12 C13 and C14 in the page header are analyzed and interpreted for selecting the correct national language option sub-sets.

30

The Teletext decoder then refers to the corresponding group of fonts in the national language option sub-set for Teletext display. When the received Teletext service is of Level 1 or 1.5, or when X/28/0 Format 1 or X/28/4 are

35

missing, the decoder will look upon the control bits C12, C13 and C14 alone for determination. In the case when the G0 character set is Latin and when C12, C13 and C14 have the value of 1, 1, 0 or 1, 1, 1, the decoder will then re-sort to the
5 single-byte language coding identification mechanism again to identify which language coding it is.

This identification mechanism will compare it to the language profiles of Czech/Slovak and Turkish, Swedish/Finnish/Hungarian and Estonian, Italian and
10 Lettish/Lithuanian and calculate the confidence value of this language coding. The language that is associated with the highest confidence value is chosen to be the final decision for the language coding. Thus, the Teletext decoder will use the corresponding supplementary set.

15 The language profile of each language is pre-evaluated based on the method presented in the cross-reference of this invention, and is installed in the decoder system. When the identification process produces exactly the same confidence value for two different languages, the system will refer to
20 other indicators to decide which language is the final solution.

According to an embodiment of the invention, there is provided a method for display of Teletext pages with an
25 identification of language coding process, the method comprising steps of:

(a) applying a single-byte language coding identification to select which character set to be used, among a plurality of character sets;

30 (b) if the character set is identified as a predetermined character set which does not have sub-character set, then displaying Teletext pages with the identified character set; and

(c) otherwise, if the character set is identified as a
35 predetermined character set which has sub-character set, then

(c1) checking a predetermined Teletext language control bits in a page header for national language option sub-set of the identified character set;

(c2) displaying Teletext pages with primary set fonts;

5 (c3) if a result indicated by the predetermined control bits to supplement the primary set fonts is clear, then

selecting the national language option sub-set fonts based on the clear result indicated by the
10 predetermined control bits to supplement the primary set fonts; and

(c4) otherwise, if a result indicated by the predetermined control bits to supplement the primary set fonts is ambiguous, then

15 (c4-1) utilizing the single-byte language coding identification process; and

(c4-2) referring to the national language option sub-set fonts based on the single-byte language coding identification result to select corresponding national
20 language option sub-set fonts for Teletext page display.

The character set may be G0 character set, the plurality of character sets may comprise Latin G0 character set, Cyrillic option 1 G0 character set, Cyrillic option 2 G0
25 character set or Cyrillic option 3 G0 character set, the predetermined character set which does not have sub-character set may be one of Cyrillic option 1 G0 character set, Cyrillic option 2 G0 character set, or Cyrillic option 3 G0 character set, the predetermined character set
30 which has sub-character set may be Latin G0 character set, the predetermined control bits may be control bits C12 C13 and C14, and the primary set fonts may be a Latin primary G0 set fonts.

35 The step (c4-1) may comprise:
determining which G0 character set to be used with

single-byte language coding identification.

The step (c4) may further comprise, before the step (c4-1):

- 5 (c4-0) determining whether to use the single-byte language coding identification when control bits produce ambiguous results when Latin G0 character set is selected.

10 In step (a) and step (c4), the single-byte language coding identification process with Teletext decoder may comprise sub-steps of:

 developing language profile for certain national languages and including the language profiles as part of Teletext decoding systems;

- 15 comparing the language coding against a potential language profiles; and

 deciding the language coding based on a confidence value of language coding.

20 The step (c4) may further comprise:

 (c4-3) developing language profiles for supported national languages.

BRIEF DESCRIPTION OF THE DRAWINGS

25

The invention, and its features and advantages will be seen from the following detailed description made with reference to the appended drawings, of which:

- 30 Figure 1 shows a flow chart of the steps of this invention, which is the detailed sequence for detecting and determining the national language option sub-set.

DESCRIPTION OF THE EMBODIMENTS

35

The steps of the invented method are described in

Figure 1. The invention is under the circumstance of Teletext Level 1 and 1.5, when the national option sub-set in use on the page is defined by control bits C12, C13 and C14 only, or of higher level of Teletext (Level 2.5 or 3.5) with missing triplet 1 in header packet X/28/0 Format 1 or X/28/4.

According to the invention, the selection of the G0 primary character set is done at the beginning, with the single-byte language coding identification mechanism. This step will decide whether Latin G0 set, or Cyrillic option 1, option 2 or option 3 will be used as the primary G0 set, based on the confidence value against the pre-set language profiles.

If the G0 character set is Cyrillic-1, or Cyrillic-2 or Cyrillic-3, the Teletext decoder will refer to the corresponding G0 set and national character sub-set for correct Teletext page display.

If the G0 set is Latin G0 character set, the selection of the set of Latin national language option characters for the display Teletext page is done by interpreting the values of control bits C12, C13 and C14 in the header packet. If the control bits are read successfully, the values will be analyzed and the Teletext decoder will designate to the corresponding Latin national language sub-set fonts based on the following designation:

C12 C13 C14 = 0 0 0	English
C12 C13 C14 = 0 0 1	German
C12 C13 C14 = 1 0 0	French
C12 C13 C14 = 1 0 1	Portuguese/Spanish

Based on the steps of the method of the invention, when the Teletext decoder reads the control bits of the above values, it will designate precisely to the corresponding Latin

national option sub-set fonts to supplement the primary G0 fonts and to display the Teletext pages correctly.

Prior to the steps described in the invention, the
5 language profiles of the Czech/Slovak vs. Turkish languages, Swedish/Finnish/Hungarian vs. Estonian, Italian vs. Lettish/Lithuanian, as well as Serbian/Croatian (for Cyrillic-1 G0 set), Russian/Bulgarian (for Cyrillic-2 G0 set) and Ukrainian (for Cyrillic-3 G0 set) are developed and
10 stored in the system based on the invention in the cross-reference of this invention.

When the G0 set is Latin one, and if the Teletext decoder reads control bits C12 C13 and C14 having the
15 values of

C12 C13 C14 = 0 1 0	Swedish/Finnish vs Estonian
C12 C13 C14 = 0 1 1	Italian vs Lettish/Lithuanian
C12 C13 C14 = 1 1 0	Czech/Slovak vs Turkish

20

According to this invention, the decoder will apply the single-byte language coding identification method stated in the cross-reference, to compare to the language profiles of respective languages and calculate the confidence value
25 against each possible option. The language with the higher confidence value will be chosen by the decoder to supplement the primary Latin G0 set, and the decoder will refer to that sub-set fonts to display Teletext pages.

What is claimed is:

1. A method for display of Teletext pages with an identification of language coding process, the method
5 comprising steps of:

(a) applying a single-byte language coding identification to select which character set to be used, among a plurality of character sets;

(b) if the character set is identified as a
10 predetermined character set which does not have sub-character set, then displaying Teletext pages with the identified character set; and

(c) otherwise, if the character set is identified as a predetermined character set which has sub-character set,
15 then

(c1) checking a predetermined Teletext language control bits in a page header for national language option sub-set of the identified character set;

(c2) displaying Teletext pages with primary set fonts;

(c3) if a result indicated by the predetermined
20 control bits to supplement the primary set fonts is clear, then

selecting the national language option sub-set fonts based on the clear result indicated by the predetermined
25 control bits to supplement the primary set fonts; and

(c4) otherwise, if a result indicated by the predetermined control bits to supplement the primary set fonts is ambiguous, then

(c4-1) utilizing the single-byte language coding
30 identification process; and

(c4-2) referring to the national language option sub-set fonts based on the single-byte language coding identification result to select corresponding national language option sub-set fonts for Teletext page display.

35

2. The method according to claim 1, wherein the

character set is G0 character set,

the plurality of character sets comprise Latin G0 character set, Cyrillic option 1 G0 character set, Cyrillic option 2 G0 character set or Cyrillic option 3 G0 character set,

the predetermined character set which does not have sub-character set is one of Cyrillic option 1 G0 character set, Cyrillic option 2 G0 character set, or Cyrillic option 3 G0 character set,

the predetermined character set which has sub-character set is Latin G0 character set,

the predetermined control bits are control bits C12 C13 and C14, and

the primary set fonts are a Latin primary G0 set fonts.

3. The method according to claim 2, wherein the step (c4-1) comprises:

determining which G0 character set to be used with the single-byte language coding identification.

4. The method according to claim 2, wherein the step (c4) further comprises, before the step (c4-1):

(c4-0) determining whether to use single-byte language coding identification when control bits produce ambiguous results when Latin G0 character set is selected.

5. The method according to claim 2, wherein, in step (a) and step (c4), the single-byte language coding identification process with Teletext decoder comprises sub-steps of:

developing language profile for certain national languages and including the language profiles as part of Teletext decoding systems;

comparing the language coding against a potential language profiles; and

deciding the language coding based on a confidence

value of language coding.

6. The method according to claim 4, wherein the step (c4) further comprises:

5 (c4-3) developing language profiles for supported national languages.