International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Report ITU-R BT.2342-0

(02/2015)

Production, emission and exchange of closed captions for all worldwide language character sets (latin and non-latin)

BT Series

Broadcasting service (television)

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Annex 1 of Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from http://www.itu.int/ITU-R/go/patents/en where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Reports

(Also available online at http://www.itu.int/publ/R-REP/en)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	**Broadcasting service (television)**
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management

***Note**: This ITU-R Report was approved in English by the Study Group under the procedure detailed in Resolution ITU-R 1.*

Electronic Publication
Geneva, 2015

REPORT ITU-R BT.2342-0

Production, emission and exchange of closed captions for all worldwide language character sets (latin and non-latin)

(2015)

Closed captions are useful for understanding the story, scene, and context in broadcast programmes by viewers with hearing difficulties but also by other people for various purposes. The broadcasting industry has utilized closed captioning systems for a long time. Originally, a closed captioning service was provided by teletext on analogue televisions, which required an additional teletext decoder to make the closed captions visible. However, the use of closed captions has been common since the closed caption feature became a standard function in digital broadcasting systems.

Thus, the number of broadcast programmes that provide closed-captioning is increasing, which helps to improve accessibility to broadcast programmes. However, the increased number of closed captioning programmes has created a new problem: the efficient creation of closed captioning data for broadcast content. Therefore, efforts have been made to develop technologies such as efficient production methods and closed captioning content exchange methods to solve, or at least mitigate, this problem.

Furthermore, the uses of closed captioning have diversified. Closed captioning is necessary for broadcasting as well at various other media such as on-demand video programmes over the Internet, and it is considered to be an important information source for advanced services. Given these new requirements, XML based closed captions standards may be solutions that provide flexibility to satisfy multiple aims and uses, including format conversion capability.

As a first step to consider the use of XML in closed captions, it is valuable to overview the techniques and the status of closed captions. Therefore, this Report describes the recent status of techniques used for the production, exchange and broadcast of closed captions. The following techniques are described in this Report.

Annex 1 – ARIB[1] closed caption standards based on enhancements of character encoding scheme

Annex 2 – SMPTE format for timed text

Annex 3 – EBU format for timed text (EBU-TT)

Annex 4 – ARIB Timed Text Markup Language (ARIB-TTML)

Annex 5 – TTML Text and Image Profiles for Internet Media Subtitles and Captions 1.0 (IMSC-1)

[1] ARIB stands for “Association of Radio Industries and Businesses”.

Annex 1

ARIB closed caption standards based on enhancements of character encoding scheme

1 Creation of closed-captioning

Creation of closed-captioning is generally taken as a part of programme production. From the viewpoint of the production process, there are two categories for how to create closed-captioning texts; live closed-captioning and offline closed-captioning. Live closed-captioning is used for live programmes, and offline closed-captioning is used for pre-recorded programmes.

1.1 Live closed-captioning

In live broadcasting programmes such as news closed-captioning text cannot be prepared prior to the time of broadcast. Live closed-captioning is used for this kind of programmes. Creation of the text is based on the vocal sound which has just been broadcasted, and the closed-captioning text is created to catch up with progress of broadcasting programmes. From technical viewpoint, minimization of the delay and accuracy of the text are the most important points.

There are several methods of live closed-captioning:

– **"Direct" speech recognition**:

  Texts are created by using direct speech recognition of announcers and reporters of the programme. It is the most efficient method to create the text, but accuracy of the created texts is not good enough for captioning in many cases due to the limitation of the speech recognizer.

– **"Re-Speak" speech recognition**:

  As described in Report ITU-R BT.2207-2 – Accessibility to broadcasting services for persons with disabilities, texts are created by using speech recognition with a captioning re-speaker. In this method, because back-ground noise can be eliminated for speech recognition, better recognition result is obtained than "Direct" speech recognition. And a re-speaker can rephrase vocal phrases of actual speakers in order to adjust the length of captioning texts.

– **"Direct typing" method**:

  In this method, captioning operators transcribe captioning texts in accordance with vocal sound of speakers by using normal keyboard or stenograph keyboard. To obtain good texts quickly, well trained operators are required.

1.2 Offline closed-captioning

In pre-recorded programmes such as soap opera, there is a time to prepare captioning texts before broadcast in many cases. In such a case, captioning operators listen to a pre-recorded video and transcribe the dialogue. Created texts and transmission time information are stored in a computer, and used at the time of broadcast.

2 Closed-captioning standard for broadcasting service

In integrated services digital broadcasting (ISDB) systems, closed-captioning scheme is defined in Vol. 1, Part 3 of ARIB STD-B24 – Data coding and transmission specification for digital broadcasting. Data is transmitted in a form of independent elementary stream. In the elementary stream, independent packetized elementary stream (PES) transmission format is used to add time synchronization function to video image.

Time stamp for closed-captioning is added to closed-captioning data, of which values are the same as presentation time stamp (PTS) of video stream.

Independent PES format for closed-captioning can convey eight languages. It also provides several presentation modes such as forced presentation and switchable presentation by the user. Attachment 1 to this Annex provides some more detailed information.

3 Exchange format of the closed-captioning file

When broadcasting programmes are exchanged, associated closed-captioning data should also be able to be exchanged. Thus, it is important to establish exchange format for closed-captioning.

ARIB STD-B36 – Exchange format of the digital closed-captioning file for digital television broadcasting system, is a standard for closed-captioning data exchange format for digital TV environment. This standard defines formats of closed-captioning text, control data, and file name convention. When using NAB[2] format file (analogue closed-captioning file for analogue television broadcasting system) in digital TV environment, a format converter to ARIB STD-B36 is required. Due to historical reasons, file name convention is designed so that the standard used for exchange can be identified by the file name. When transmitting by serial digital interface (SDI), closed-captioning data formatted by this standard is embedded in ancillary data area, as defined in ARIB STD-B37 – Structure and operation of closed-captioning data conveyed by ancillary data packets. When transmitting by MPEG2 transport stream, ARIB STD-B24 is used as similar to broadcasting. For more details, see Attachments 2 and 3 for ARIB STD-B36 and STD-B37, respectively.

4 Use of captioning data of broadcasting content on the Internet

Recently, broadcasting programmes are also offered on the Internet based environment as an on-demand service. On-demand video clips may be watched by various devices including PC and TV. Service functionality of video on-demand should be the same on any devices, thus closed-captioning should be available even on the web. An efficient way to provide closed-captioning on the video on-demand content is to convert closed-captioning data from the broadcasting format. For this purpose, a Format converter has been developed jointly by NHK and LSI JAPAN Co. Ltd. An output format of this converter is based on the bundled functionality on closed-captioning playback of the typical player software on the web. Extension of the player software has been developed as well so that functionalities of the player software on closed-captioning are equivalent to those on broadcasting including Ruby[3] and Gaiji[4].

---

[2] NAB stands for "The National Association of commercial Broadcasters in Japan". The name of Association was changed to JBA (The Japan commercial Broadcasters Association) in 2012.

[3] Ruby is a set of small-sized kana characters attached above Kanji character to explain how to read it.

[4] Gaiji is ideographical Kanji characters which are not supported by the system.

By using these converter and player software, the same service can be provided on both broadcasting and video on-demand.

To expand availability of format conversion, General Caption Markup Language (GCML) is also developed. GCML is a markup language and works as an intermediate format for format conversion across various media including broadcasting and the web. When a new format is required for some purposes, only a converter between the new format and GCML is needed for re-use of the created closed-captioning data for the purposes.

For the purpose of realizing caption production by lower cost, simplified caption markup language is under consideration at ARIB.

5 Consideration on efficient use of closed-captioning data on various media

The needs of closed-captioning are increasing day-by-day. And as described in previous section, use of closed-captioning is not limited to broadcasting. It is convenient and efficient that created closed-captioning data can be re-used in different media and different environment. In this section, possibility of re-use of closed-captioning data is discussed.

5.1 Language specific requirements

Considering that broadcasting programmes are exchanged internationally, the mechanism for re-use of closed-captioning data should satisfy language specific requirements. For example, in case of Japanese language, followings are language specific requirements.

– **Ruby**:

  Ruby is supplement information on how to read Kanji character. Some Kanji characters are not common for typical Japanese persons and such Kanji requires ruby information. And in case of name of persons or places, even common Kanji characters should be read in special readings in some cases. Ruby is very convenient in such cases as well.

– **Gaiji**:

  Computers and broadcasting equipment such as TV implement standardized set of Kanji characters. Typical TV receivers have about 6 400 Kanji characters font set. In some cases, Kanji characters not supported by the system are required to describe information. Gaiji is a mechanism to render such unsupported Kanji characters on the devices.

– **Vertical writing**:

  Historically, Japanese language is vertical writing. Although it is common to write Japanese in horizontal writing today, there is a need to write closed-captioning vertically in some cases.

It is believed that there are more language specific requirements for a mechanism for re-use of closed-captioning internationally. Consideration on support of those language specific requirements is required not to drop the information in closed-captioning or intention of closed-captioning creators.

5.2 Format conversion of closed-captioning data

In digital broadcasting systems, each system specification provides its own closed-captioning standard. Considering compatibility with various types of existing digital broadcasting receivers, format conversion at production side may be an efficient approach for re-use of closed-captioning data. In this case, an intermediate format like GCML may be one of the practical approaches. A well-designed intermediate format will have capability to convert closed-captioning data to and from a wide range of media, including web, blu-ray, and broadcasting.

Attachment 1
to Annex 1

ARIB STD-B24

Data coding and transmission specification for digital broadcasting

1 Scope

The standard is applied to data transmission for data broadcasting carried out as part of digital broadcasting.

2 Presentation function of caption and superimpose

Presentation function of the caption is shown in Table A1-1.

TABLE A1-1

Presentation function of caption

Display function	Format	1920 x 1080, 960 x 540, 1280 x 720, 720 x 480 (each of them is mixed with vertical and horizontal writing format)
	Character set	Kanji, hiragana, katakana, symbol, alphanumerical, Greece characters, Russian characters, ruled line, DRCS
	Font	Plural typeface can be designated
	Supplemental Characters (Gaiji)	By DRCS graphics
	Character display size	Size designation and deformation in pixel unit, standard, 1 x 2, 2 x 1, 2 x 2, 1/2 x 1, and 1/2 x 1/2 are directly designated using control code.
	Coloring	256 colors are displayed simultaneously (color map used, output: color value of YCBCR and α value (8-bit x 4))
	Character coloring unit	Each character (outer frame of character or character display block)
	Character attribute	Reversing polarity, flashing, underline, enclosure, shading, bold, italic, bold and italic
	Graphics	Geometric, bitmap
Display control	Timing control	Display timing, erase timing
	Switching control	Cut, dissolve, wipe, slide, and roll
Others	Language	up to 8 languages per 1 ES
	Music data	For coding synthesized sound, coding method shall be in accordance with standard method of transmission related to television superimpose broadcasting (ARIB STD-B5).
	ROM sound	PCM (AIFF-C)

3 Independent PES transmission protocol

The independent PES transmission protocol is a method used to implement streaming for data broadcasting services. The independent PES transmission protocol defined in this section has two types: synchronized type and asynchronous type.

The synchronized PES transmission protocol is used when it is necessary to synchronize data in a stream with other streams including video and audio. The asynchronous PES transmission protocol is used when the synchronization is not necessary. As a major application example, it is expected that the synchronized type is used for transmitting captions.

Attachment 2
to Annex 1

ARIB STD-B36

Exchange format of the digital closed caption file for digital television broadcasting system

1 Scope

The standard is applied to (exchanging format of) the digital closed caption file for digital television broadcasting system.

2 Recording format of the closed caption data

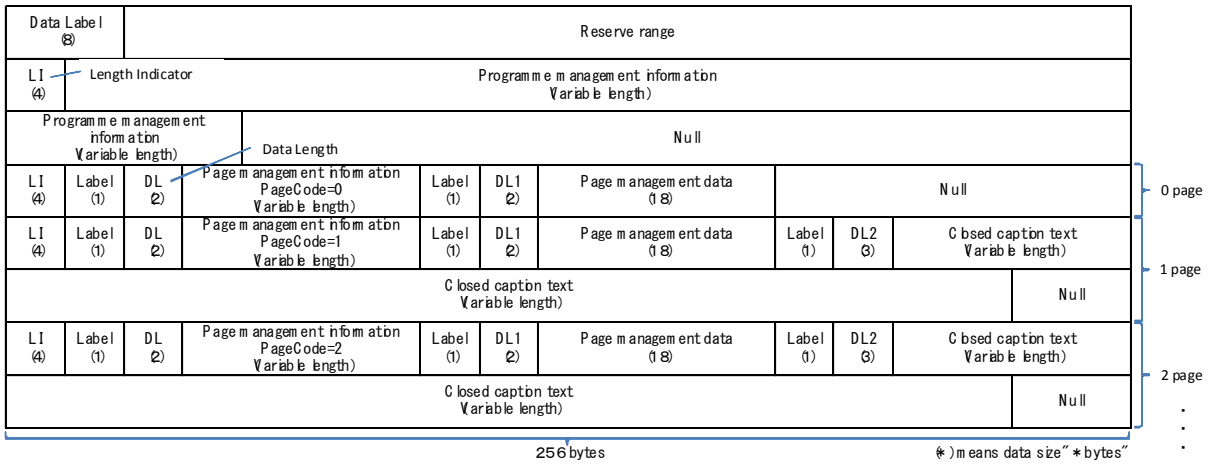
2.1 Closed caption data recording format

Each digital closed caption files are recorded in the same directory for every programme. Digital closed caption file consists of programme control information, page management information, caption management data and caption statement data.

Figure A2-1 illustrates the arrangement and internal makeup of closed caption data.

FIGURE A2-1

Arrangement and internal make-up of closed caption data



– **Programme management information**:

Programme management information consists of programme information indicating programme title, programme schedule, and so on.

– **Page management information**:

Page management information consists of page information indicating carried out timing, delete timing, font size, video aspect, presence or absence of rollup, and so on.

– **Page management data**:

Page management data consists of page management data header indicating language or transmission mode of the caption and zero or more than one data unit, following it.

Attachment 3
to Annex 1

ARIB STD-B37

Structure and operation of closed caption data conveyed by ancillary data packets

1 Scope

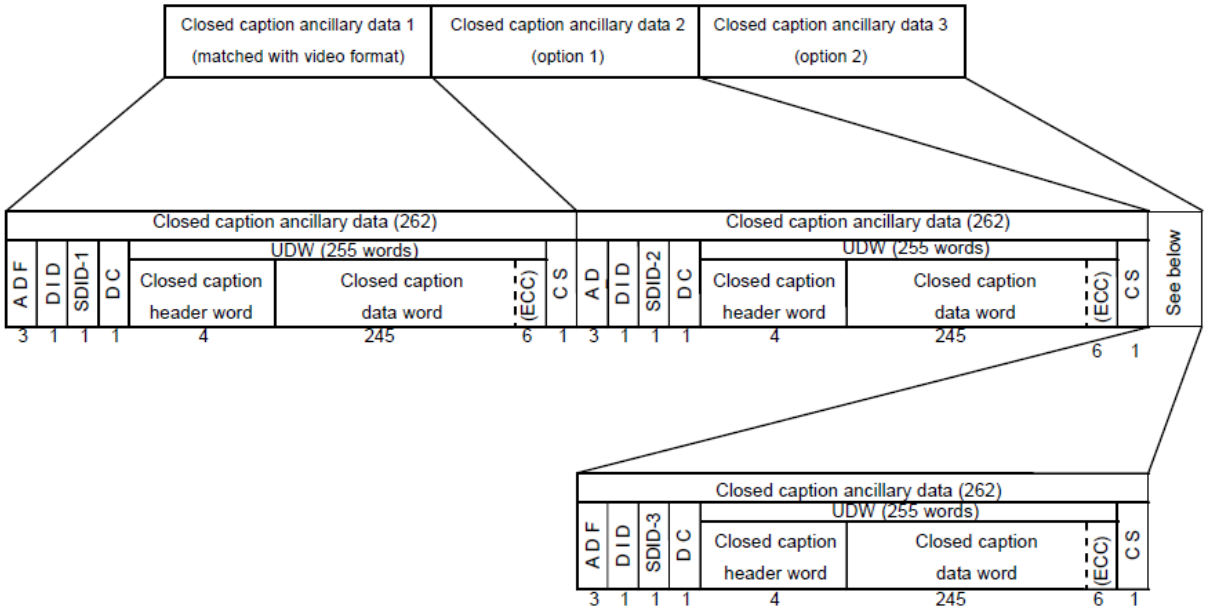
This standard is applicable to devices which convey closed caption data using ancillary data packets.

2 Packet structure and arrangement

Figure A3-1 illustrates the arrangement and internal makeup of closed caption ancillary data packets which are embedded in one line of video. In this structure, one word comprises 10 bits.

FIGURE A3-1

Arrangement and internal make-up of closed caption ancillary data packets



Annex 2

SMPTE format for timed text

The Society of Motion Picture and Television Engineers (SMPTE) has been working for some time on standards for carrying text that is designed to accompany a television programme, but delivered separately from, or in addition to, the image signal. These may be used for subtitles, closed captions and subtitles for the deaf and hard of hearing (SDH). The system can also be used for other applications where timed text needs to be synchronized with an image.

The group of standards, which is termed SMPTE-TT, defines a file format which extends and constrains W3C Timed Text Markup Language (TTML). The SMPTE-TT file includes information on the text itself as well as positioning and timing information. It thus provides the means of conveying captions or subtitles in virtually any environment such as in program production or delivery via an IP network (managed or unmanaged networks).

Recent standardization work on SMPTE-TT has been to clarify the conversion from CEA 608 (US analog closed captioning) and the creation of a new conversion from CEA 708 (US digital closed captioning). We believe this may be relevant to closed captions referred to in Annex 2 of Recommendation ITU-R BT.1301-2.

Below is a list of links to the relevant SMPTE-TT documents, which are available for download at no charge.

– General overview of document structure ST 2052-0:2013
– Standard document (Definition of SMPTE-TT) ST 2052-1:2013
– Recommended Practice (Conversion from CEA 608 to SMPTE-TT) RP 2052-10:2013
– Recommended Practice (Conversion from CEA 708 to SMPTE-TT) RP 2052-11:2013
– FAQ

Annex 3

EBU format for timed text (EBU-TT)

The EBU is a professional association of public service broadcasters in 56 countries. It has been developing a means of handling timed text that is generated for the purposes of facilitating the understanding of television programmes of those with impaired hearing or for providing foreign language subtitling.

Mindful that WP 6B is working towards a "Preliminary Draft New Report on Production, Emission, and Exchange of Closed Caption for all worldwide language character sets (latin and non-latin)", the EBU would like to draw the attention of the Working Party to several EBU specifications that are pertinent to this new Report.

Below is a list of links to EBU publications that form part of a family of EBU-TT specifications that together cover the chain from production to distribution, taking into account professional users' requirements in terms of operational efficiency, interoperability, archiving/repurposing and (automatic) translation to other – e.g. platform-specific – subtitling formats.

[1] is the specification of EBU STL, a subtitling exchange format that is widely used in the industry for the exchange and archiving of subtitles. A very large base (probably millions) of STL files exists in archives around the world.

[2] is the main specification of EBU-TT. The format is based on W3C TTML 1 and it was developed in close coordination with users and manufacturers.

[3] is the specification of EBU-TT-D, a delivery specification for use over IP networks, including but not limited to use as the subtitling standard for DVB Dash (Blue Book A168) and the upcoming HbbTV 2.0. EBU-TT-D is a highly constrained subset of W3C TTML 1, with two extensions to improve rendering of subtitles in specific cases.

[4] specifies how EBU-TT-D information can be stored using the storage format of the ISO base media file format (ISOBMFF) defined in ISO/IEC 14496-1.

[1] https://tech.ebu.ch/docs/tech/tech3264.pdf

3264_e.doc

[2] https://tech.ebu.ch/docs/tech/tech3350.pdf

tech3350_v1 0.doc

[3] https://tech.ebu.ch/docs/tech/tech3380.pdf

tech3380-final.doc

[4] https://tech.ebu.ch/docs/tech/tech3381.pdf

tech3381-final.doc

A further EBU document concerning the mapping from EBU STL to EBU-TT is currently under review (v0.9 published as https://tech.ebu.ch/docs/tech/tech3360.pdf). When the final version becomes available, it will be communicated to WP 6B.

Annex 4

ARIB Timed Text Markup Language (ARIB-TTML)

1 Introduction

ARIB-TTML is an extension of the second edition of the TTML1 specification defined by the World Wide Web Consortium (W3C) and SMPTE ST 2052-1:2013 "Timed Text Format (SMPTE-TT)". Considering the requirements for broadcasting content, the presentation capability of ARIB-TTML should be enhanced to match that of ARIB STD-B24, which is used at present in Japan. It is important that ARIB-TTML has at least the same presentation capability as the existing closed caption standard to ensure the content quality and to facilitate the transition to the new standard. That is, ARIB-TTML should satisfy language specific requirements for Japanese described in § 5.1 in Annex 1. In addition, the new closed caption standard is expected to be applicable to broadband delivered content, which may use a different presentation environment compared with TV receivers. Based on these considerations, ARIB-TTML provides many enhancements relative to TTML1. For example, ARIB-TTML is capable of handling pictures using the method defined in SMPTE ST 2052-1:2013. Furthermore, ARIB-TTML has the capability to handle audio clips. This Annex provides information on this extension and it considers the actual use of XML based closed captions considering ARIB STD-B62.

ARIB STD-B62 can be found at:

http://www.arib.or.jp/tyosakenkyu/kikaku_hoso/hoso_kikaku_number.html

2 Presentation capability

Presentation capability of ARIB-TTML is enhanced compared with that of ARIB STD-B24. Table 4.1 summarizes the presentation capability of ARIB-TTML.

TABLE 4.1

Presentation capability of ARIB-TTML

Presentation function	Format	1 920 × 1 080, 3 840 × 2 160, 7 680 × 4 320
	Orientation	horizontal, vertical, mixture of horizontal and vertical
	Character repertoire	JIS X0201:1997, JIS X0213:2004, Latin-1 supplement of ISO/IEC 10646:2012, as well as additional Kanji and the symbols defined in Vol. 1 Part 2 of ARIB STD-B62
	Character encoding	UTF-8
	Gaiji	Available by SVG 1.1 and WOFF File Format 1.0
	Font	Selectable
	Font size	Selectable in pixels
	Font colour	256 steps each for red, green, blue and alpha (transparency)
	Colour designation unit	Each character
	Character attribute	flashing, underline, overline, strike-out, enclosure, shadow, bold, italic, italic bold
	Graphics	PNG, SVG
Presentation control	Timing control	start time, end time, duration
	Switching control	cut, pop-on, paint-on, roll-up, scroll, key-frame animation
Misc.	Audio clip	PCM (AIFF-C), MP3, MPEG2-AAC, MPEG4-AAC
	Built-in sound	PCM (AIFF-C), MP3, MPEG2-AAC, MPEG4-AAC

3 Enhancement in ARIB STD-B62

3.1 Functional enhancement

To achieve presentation capability listed in Table 4.1, following enhancements are defined in ARIB STD-B62.

1) Animation

Animation is presented based on combinations of the following elements and attributes.

arib-tt:keyframes: This element is the root element of an animation sequence. Each step in an animation is defined by the arib-tt:keyframe element. Each animation sequence can be distinguished by the name of the animation, which is given by the “animationName” attribute of this element.

arib-tt:keyframe: This element represents each step of animation by specifying position, colour, font, opacity, etc.

arib-tt:animation: This attribute is used to refer to a sequence in an animation that is defined by the arib-tt:keyframes element. This attribute can also specify the duration, timing-function, delay, iteration count and direction. This attribute can be used in the style, body, div, p, and span elements.

These elements and attributes can also be used for switching control.

arib-tt:marquee: This attribute describes an automated scrolling movement by specifying the style, direction, speed, and play count. The style is one of the following: scroll, slide, and alternate. This attribute can be used in the style, body, div, p, and span elements.

2) Additional font handling

arib-tt:font-face: This element allows the use of a font or font family that is not preinstalled in a receiver including Gaiji. This element can be a child element(s) of the styling element.

3) Ruby

In some cases, it is necessary to use Ruby to read Kanji characters. In ARIB STD-B24, this is achieved by allocating small Hiragana or Katakana characters along with a targeted Kanji character. In ARIB STD-B62, a dedicated attribute is defined to make a processing system recognize the existence of Ruby and the content.

arib-tt:ruby: This attribute specifies and describes a target of ruby and the content of ruby. This attribute can be used in the elements of div, p, and span.

4) Character decoration and presentation control

The following attributes are defined in ARIB STD-B62 for character decoration. These attributes add character decoration functions to CSS functions in TTML 1.0.

arib-tt:border: This attribute describes the style, thickness, and colour of the border lines used to enclose characters. The style can be one of the following: none, hidden, solid, double, groove, ridge, inset, outset, dashed, and dotted. This attribute can be used in the style, body, div, p, and span elements.

arib-tt:letter-spacing: This attribute describes the character spacing for rendering as pixels. This attribute can be used in the style, body, div, p, and span elements.

arib-tt:text-shadow: This attribute describes the control of shadows for rendering. The available controls are offset-x, offset-y, blur-radius, and colour. Offset-x and offset-y represent the horizontal and vertical offsets of the shadow’s position in pixels. Blur-radius represents the gradation radius in pixels.

5) Image

The “image” element and “backgroundImage” attribute defined in SMPTE ST 2052-1:2013 can be used to present an image.

6) Audio

Audio playback is supported in ARIB-TTML. However, because this is a feature within a closed caption standard, it is assumed that the audio playback comprises a short sound such as chime.

arib-tt:audio: This element represents an audio clip file. The location of an audio clip file and the control required for repeated playback are given by the attributes of this element.

3.2 Name space

ARIB TTML inherits characteristics of XML document. Thus, the name space definition is important for identifying elements and attributes. Table 4.2 lists the name spaces and their prefixes.

TABLE 4.2

Name spaces in ARIB-TTML

Name	Prefix	Name space
TT	tt:	http://www.w3.org/ns/ttml
TT Parameter	ttp:	http://www.w3.org/ns/ttml#parameter
TT Style	tts:	http://www.w3.org/ns/ttml#styling
TT Metadata	ttm:	http://www.w3.org/ns/ttml#metadata
SMPTE	smpte:	http://www.smpte-ra.org/schemas/2052-1/2013/smpte-tt
ARIB	arib-tt:	http://www.arib.or.jp/ns/arib-ttml/v1_0

4 Transmission of closed captions

An ARIB-TTML document is designed to contain closed caption data in a single language. If closed captions are offered in multiple languages, each set of closed caption data, including the images and audio (if any), is transmitted separately, e.g. by different modules of the DSM-CC carousel or by different MMT assets.

The following three methods can be considered for the transmission of each closed caption dataset.

1) Transmission of a complete document repeatedly. In ARIB-TTML, this is known as program mode.

2) Transmission of a document containing fragmented closed caption data for a programme. In ARIB-TTML, this is known as segment mode.

3) Transmission of a frequently updated document containing fragmented closed caption data. In ARIB-TTML, this is known as live mode.

In each method, a document should contain appropriate descriptions according to the transmission method used because the description form is strongly related to the timing of closed caption decoder initialization in a receiver. A closed caption decoder is initialized in a receiver when a new or updated document is received. After initialization, all of the data and the presented captions are erased. It is also necessary to define a method for delivering external entities such as image files and audio clip files in a broadcast signal, and to refer to them from an ARIB-TTML document.

4.1 Program mode

The program mode is used to transmit closed caption data with an ARIB-TTML document that contains all of the closed caption data for a programme. Because viewers may tune into the program after it has started, the same ARIB-TTML document is transmitted repeatedly throughout the programme.

During transmission in the non-live mode, the closed caption decoder in a receiver is responsible for initializing the screen, audio playback, image presentation, and Gaiji loading when an ARIB-TTML document updates. This is sufficient for the static representation of closed caption data, such as an ARIB-TTML document that contains all the closed caption data for a broadcast programme. Thus, this mode assumes that an ARIB-TTML document will not be updated during the corresponding broadcast programme.

4.2 Segment mode

The program mode is an easy way to provide closed captioning for a programme. However, this mode requires a specific amount of memory in a receiver to process the data and a wide bandwidth for transmissions. The segment mode is a transmission method that reduces these resource requirements. In the segment mode, an ARIB-TTML document contains only the necessary closed caption data for a specific moment in a programme. By transmitting the ARIB-TTML document with fragmented closed caption data once for each document, the correct caption will appear on the screen when required because initialization of the closed caption decoder occurs every time the documents are received. In segment mode, the documents are complete ARIB-TTML documents and each document does not depend on other documents. The "begin" and "end" attributes in the document allow the captions to be presented at the appropriate times.

4.3 Live mode

The program and segment modes handle static closed captions, i.e. the closed caption data are determined before the time of the programme. By contrast, the live mode can be used if dependencies exist among the ARIB-TTML documents due to high frequency updates in the segment mode, or if the caption presentation time is not predetermined for the programme. In particular, the live mode is suitable for live programmes.

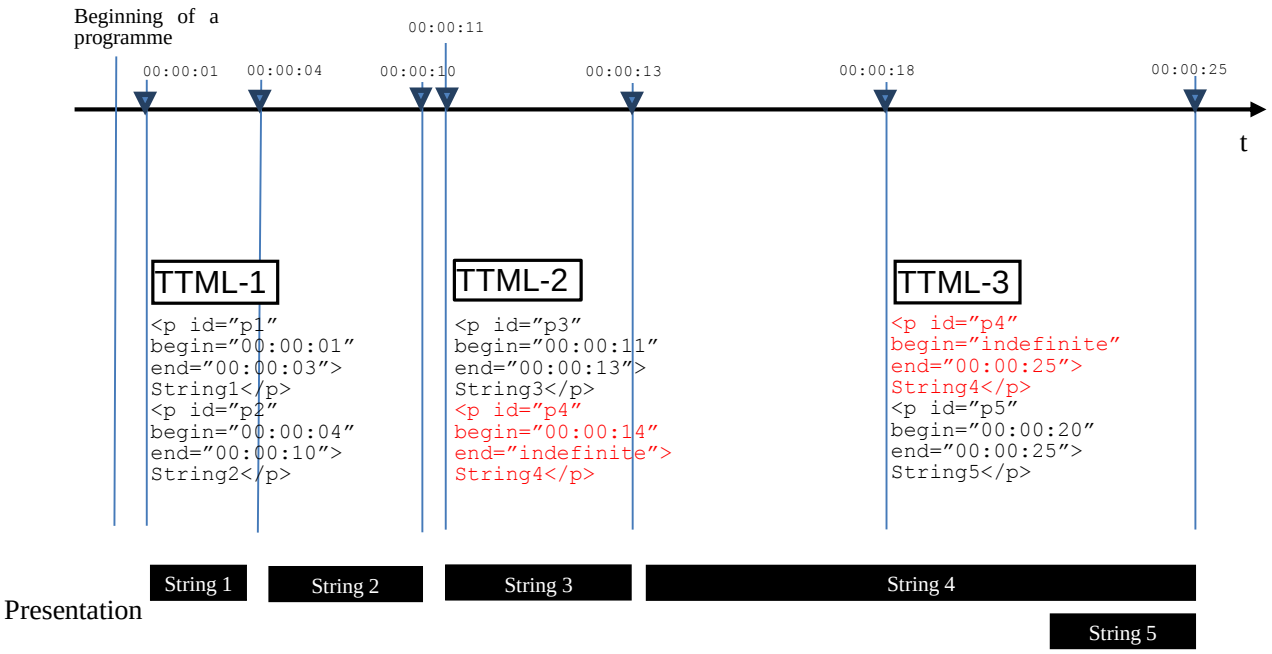
In live mode, closed caption decoder initialization is dependent on conditions. Thus, a closed caption decoder is not initialized when all of the following conditions are satisfied.

1) A value of "indefinite" is set for the "end" or "dur" attribute of an element in an ARIB-TTML document.
2) The "id" attribute is also set for the element.
3) In a consecutive ARIB-TTML document, an element with the "id" attribute has the same value of 2 and the value of the "begin" attribute is "indefinite".

This process allows the continuous presentation of a specified closed caption across multiple ARIB-TTML documents. An example of this process is shown in Fig. 4.1.

FIGURE 4.1

Example of live mode transmission using ARIB-TTML



Strings 1, 2, and 3 are presented according to the values of the "begin" and "end" attributes. String 4 appears in both the TTML-2 and TTML-3 documents. TTML-2 document specifies values for id = "p4" and end = "indefinite". In the TTML-3 document, the same id is given to this string, which indicates that the "begin" attribute is "indefinite". The conditions given above are satisfied by these descriptions, which means that the closed caption decoder is not initialized for the TTML-3 document. Thus, String 4 is displayed continuously on the screen until the time indicated by the "end" attribute in the TTML-3 document.

In the live mode, it is not possible to use the "begin", "end", and "dur" attributes. In this case, a closed caption decoder will render the closed caption data immediately and hold them. To erase the closed caption data from the screen, it is necessary to transmit a "null" ARIB-TTML document that only comprises the root element.

4.4 External entity reference

When transmitting image files or audio clip files during a broadcast signal, it is necessary to define how to refer them from an ARIB-TTML document. In ARIB STD-B60 – Media transport using MMT in digital broadcasting or ARIB STD-B24 (for delivery using MPEG2-TS), the external entities are packed into the same container with an ARIB-TTML document. In a container, each entity is referred as a "subsample". An ARIB-TTML document is always placed as the first entity in a container, and thus it can be referred to as "subsample 0". The other entities follow the ARIB-TTML document in a container and they are referred to by their corresponding subsample numbers in order. In the ARIB-TTML document description, these entities are referred to by using the prefix "subt://" followed by the subsample number.

Reference to subsamples within the same container is valid and it occurs before the reception of the next transmission unit.

5 How to satisfy Japanese language specific requirements

As described in § 5.1 in Annex 1, there are three language specific requirements to Japanese language. Ruby and Gaiji are satisfied as described in § 3.2 in this Annex. Vertical writing is satisfied by one of the attributes of the style element (tts:writingMode) defined in TTML1.

Annex 5

TTML Text and Image Profiles for Internet Media Subtitles and Captions 1.0 (IMSC-1)

1 Summary

TTML Text and Image Profiles for Internet Media Subtitles and Captions 1.0 (IMSC1) is a Candidate Recommendation of World Wide Web Consortium (W3C) published on December 9, 2014. The Candidate Recommendation can be found at:

http://www.w3.org/TR/2014/CR-ttml-imsc1-20141209/

IMSC1 specifies two profiles of [TTML1]: a text-only profile and an image-only profile. These profiles are intended to be used across subtitle and caption delivery applications worldwide, thereby simplifying interoperability, consistent rendering and conversion to other subtitling and captioning formats. The text profile is a superset of [SDPUS]. The both profiles are based on [SUBM].

The specification defines extensions to [TTML1], as well as incorporates the extensions specified in [ST2052-1] and [EBU-TT-D].

The information contained in this Annex is provided by communication between W3C and ITU-R.

2 Motivation of the development and the scope of the standard

W3C TTWG is chartered [1] to simultaneously work on TTML and WebVTT, with the intention that both will be published as Recommendations according to the W3C Process [2], and to publish a mapping between them as a working group Note (§ 1.3).

[1] http://www.w3.org/2014/03/timed-text-charter.html

[2] http://www.w3.org/2014/Process-20140801/

The most recent published version of IMSC1 is the Candidate Recommendation, published 9th December 2014.

The most recent published version of WebVTT is the First Public Working Draft, published 13th November 2014 [3].

[3] http://www.w3.org/TR/2014/WD-webvtt1-20141113/

3 Progressive decode

The IMSC1 progressivelyDecodable flag signals that the presentation of an IMSC1 document can proceed before the document is received in its entirety, i.e. the presentation at time T1 does not depend on the presentation at time T2, where T1 < T2.

The IMSC1 progressivelyDecodable flag however assumes that the parsing of the IMSC1 document started at its beginning, and is thus not intended to allow the parsing (and presentation) of an IMSC1 document to start at an arbitrary location within: information contained in the <head> element is generally required for proper presentation.

To allow viewers to start watching a TV programme after it started, the subtitle/caption essence can be split into successive standalone IMSC1 progressivelyDecodable documents, each occupying a non-overlapping interval of the program timeline. The IMSC1 presentation can then start as a soon as the beginning of such an interval is encountered.

4 Precise time expression and synchronization

IMSC1 is designed to support high frame rate video images.

Clause 6.6 allows the author to accurately time a subtitle/caption so that it is displayed on a specific frame of the related video object used for authoring. This is particularly important around scene boundaries. Clause 6.6 is not intended to constrain display refresh rate and related video object frame rate.

For instance:

– assuming a caption/subtitle essence is authored against 120 fps video;

– a scene transition occurs at the boundary between frames #725 and #726;

– a caption/subtitle X is intended to be removed before this scene transition.

The author could specify end="6.04s" for caption/subtitle X, ensuring that caption/subtitle X is displayed on frame #725, but not frame #726.

Clause 9 specifies a Hypothetical Render Model (HRM), which is used to constrain document complexity by effectively limiting the rate and size of characters or images to be displayed over time. HRM processing is specified independently of display refresh rate and related video object frame rate. Furthermore, HRM limits are intended to accommodate common subtitle/caption practice, which are driven by human factors such as reading rates, and are thus independent of display refresh rate and related video object frame rate.

In conclusion, neither clause 6.6 nor clause 9 appear to limit the application of IMSC1 to related video object with high frame rate video images up to 120 Hz, as defined in Recommendation ITU-R BT.2020.

5 Designation of Text and Image profiles

IMSC1 is designed to address the following scenarios:

a) the subtitle/caption essence can be represented as text in its entirety. In this scenario, a single document conforming to the Text profile is offered. This is the preferred scenario;

b) it is not possible (or desirable) to represent the subtitle/caption essence as text in its entirety, e.g. all elements cannot be reliably represented using a combination of Unicode codepoints, commonly available fonts and TTML1 styling. In this scenario, a document conforming to the Image profile is offered. In addition:

i) text metadata can be associated with individual images to support content indexing and facilitate quality checking of the document during authoring (see § 6.7.4);

ii) a separate standalone document conforming to the Text profile can be simultaneously offered, if required/desired.

The IMSC1 approach, where Text and Image profiles are mutually exclusive, was selected for simplicity and reliability: if text and image essence were mixed in one document, it would be necessary to specify the relationship between text essence and related image essence, and the rules for rendering one instead of the other without compromising layout. Such a model is not defined in either TTML1 or SMPTE-TT. Note that there is, in general, no one-to-one correspondence between text and image essence, e.g. placement and timing.

This approach is consistent with:

– current practice where image and text subtitle/captions are treated separately, c.f. D-Cinema, blu-ray, etc.;

– the widespread use of multiplexed containers, e.g. ISOBMFF and MXF, that allow multiple video, audio and subtitle/caption tracks to be associated with a common timeline and delivered together.

W3C TTWG is actively working on TTML2 (see the latest Editor's Draft at [4]) and considering future versions of IMSC1.

[4] https://dvcs.w3.org/hg/ttml/raw-file/tip/ttml2/spec/ttml2.html