INTERNATIONAL STANDARD

ISO/IEC 14496-17

First edition
2006-04-01

Information technology — Coding of audio-visual objects —

Part 17: Streaming text format

Technologies de l'information — Codage des objets audiovisuels — Partie 17: Format de texte en flux

Reference number
ISO/IEC 14496-17:2006(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC 14496-17 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

ISO/IEC 14496 consists of the following parts, under the general title *Information technology — Coding of audio-visual objects*:

— *Part 1: Systems*

— *Part 2: Visual*

— *Part 3: Audio*

— *Part 4: Conformance testing*

— *Part 5: Reference software*

— *Part 6: Delivery Multimedia Integration Framework (DMIF)*

— *Part 7: Optimized reference software for coding of audio-visual objects* [Technical Report]

— *Part 8: Carriage of ISO/IEC 14496 contents over IP networks*

— *Part 9: Reference hardware description* [Technical Report]

— *Part 10: Advanced Video Coding (AVC)*

— *Part 11: Scene description and application engine*

— *Part 12: ISO base media file format*

— *Part 13: Intellectual Property Management and Protection (IPMP) extensions*

— *Part 14: MP4 file format*

— *Part 15: Advanced Video Coding (AVC) file format*

— *Part 16: Animation Framework eXtension (AFX)*

— *Part 17: Streaming text format*

— *Part 18: Font compression and streaming*

— *Part 19: Synthesized texture stream*

— *Part 20: Lightweight Application Scene Representation (LASeR) and Simple Aggregation Format (SAF)*

— *Part 21: MPEG-J GFX*

— *Part 22: Open Font Format*

Introduction

This International Standard was developed in response to the need for a generic method for coding of text at very low bitrate as one of the multimedia components within audiovisual presentations. This International Standard allows for example subtitles and Karaoke song texts to be coded and transported as separate text streams for presentation jointly with other components of an audiovisual presentation at bitrates that are sufficently low for use in mobile services over IP.

Information technology — Coding of audio-visual objects —

Part 17:
Streaming text format

1 Scope

This International Standard specifies the coded representation of textual information for timed presentation on screens. The text may be streamed in association with video and audio, in which case the text may represent subtitles e.g. with translations of the associated audio in another language, or as an aid to the hard of hearing; another example is the text of a song in a Karaoke application. However, the text may also be streamed as a stand-alone application without any associated video and audio. The streaming text format is specified in a transport agnostic manner, so as to allow transport over a large variety of transport means, while providing a reasonable level of random access and error robustness.

The text streams are defined as byte streams that are capable of carrying text access units of a specified format, optionally interleaved with data needed for the decoding of the text stream. The format of text streams and text access units is specified, as well as signaling and decoding of text streams.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 14496-18:2004, *Information technology — Coding of audio-visual objects — Part 18: Font compression and streaming*

3GPP TS 26.245: 2003, Timed text format (Release 6)

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

3.1
text stream
byte stream capable of carrying text access units of a specified format, optionally interleaved with data needed for the decoding of the text stream

3.2
text access unit
individually accessible portion of data within a text stream

NOTE Each text access unit contains the coded representation of text data. For presentation, the text access unit can be associated with a single time stamp.

3.3
3GPP text stream
text stream carrying 3GPP text access units

3.4
3GPP text access unit
text access unit carrying data from a text sample specified by 3GPP

3.5
text sample
when used in the context of a 3GPP text stream, a text sample, as specified in 3GPP TS 26.245, consisting of a text string, optionally followed by one or more text modifiers

3.6
text string
when used in the context of a 3GPP text stream, data within a text sample, representing a string of characters encoded using UTF-8 or UTF-16, as specified in 3GPP TS 26.245

3.7
text modifier
when used in the context of a 3GPP text stream, data within a text sample, specifying a modification to the presentation of the text string within that text sample, as specified in 3GPP TS 26.245

3.8
sample description
when used in the context of a 3GPP text stream, descriptive text data, providing global information about one or more text samples, such as font(s) to be used and positioning of the text, as specified in 3GPP TS 26.245

3.9
Timed Text Unit
TTU

syntactical structure within a 3GPP text stream for carriage of text access units, whereby its index *j* identifies which type of data (such as a complete text access unit, a fragment thereof or a sample description) is carried

4 Text stream format

A text stream is a byte stream that is capable of carrying text access units of a specified format, optionally interleaved with data needed for the decoding of the text stream. The text stream format is defined so as to conveniently allow for carriage in transport packets and files. However, such carriage itself depends on specific formats for file storage and transport and is beyond the scope of this standard.

5 Text access units

5.1 Timing and decoding of text access units

Each text access unit contains text data of a specified format; to each text access unit a single time stamp applies. The time stamp assigned to a text access unit indicates the time at which the text access unit is to be presented on the display. To decode a text stream, a receiver needs information on the text stream, as defined by the so-called decoder configuration in TextConfig. The TextConfig signals the format of the text data and may provide information specific to the format of the text data.

5.2 Format of text access units

```
text access unit{
   textData;            // the format of the textData is textFormat specific
   }
```

5.3 TextConfig

```
TextConfig(){
   bit(8)  textFormat;
   bit(16) textConfigLength;
   formatSpecificTextConfig();
   }
```

5.4 Semantics

`textFormat` – one byte signaling the format of the text data. The value 0x01 signals that the text data carries timed text data as defined in 3GPP TS 26.245, in a manner defined in clause 7.

Table 1 — textFormat

0x00	Reserved
0x01	Timed Text as specified in 3GPP TS 26.245
0x02 – 0xEF	Reserved
0xF0 – 0xFE	User-private
0xFF	Reserved

`textConfigLength` – unsigned integer that specifies the size in number of bytes of `formatSpecificTextConfig()`.

6 Usage of a text stream within an MPEG-4 system context

6.1 Signaling of a text stream

When used in an MPEG-4 system context, a text stream shall be signaled by a streamType value 0x0D and by an objectTypeIndication value of 0x08.

6.2 Usage in the scene description

When used within an MPEG-4 Scene description, a text stream object is used as follows:

- If the text stream object is used by an AnimationStream node, it shall be presented in the scene, regardless of whether the AnimationStream node is active or not;
- If the text stream object is not used by any AnimationStream node, it shall not be presented in the scene.

Spatial presentation of the text data is specified by the underlying streaming text format. In case the scene description has display size indication, results are undefined if the positioning of text data covers an out-of-display area.

All rules regarding time control and segmentation of stream objects apply to the text stream object.

7 Text data format for 3GPP text streams

7.1 Introduction

This clause specifies the text data format for 3GPP text streams.

7.2 Carriage of text samples and sample descriptions in 3GPP text access units

3GPP TS 26.245 defines timed text data to consist of text samples and sample descriptions, and that each text sample consists of one text string, optionally followed by one or more text modifiers. Each text string represents the characters that form the text to be displayed, while the text modifiers carry the changes that are to be applied to the text string during the time that the text is to be displayed within a text box, such as text colour changes synchronized with a song for a Karaoke application.

A sample description provides global information about a text sample, for example about font(s) to be used, about the positioning of the text within the text box, the background colour of that text box, etc. Multiple sample descriptions are allowed; to each sample description (SD) an index is assigned and to each text sample the index of the applicable sample description is associated. While a sample description will typically apply to multiple text samples, to each text sample exactly one sample description applies.

The relationship between sample descriptions, text samples, a text string and text modifiers is depicted in Figure 1.

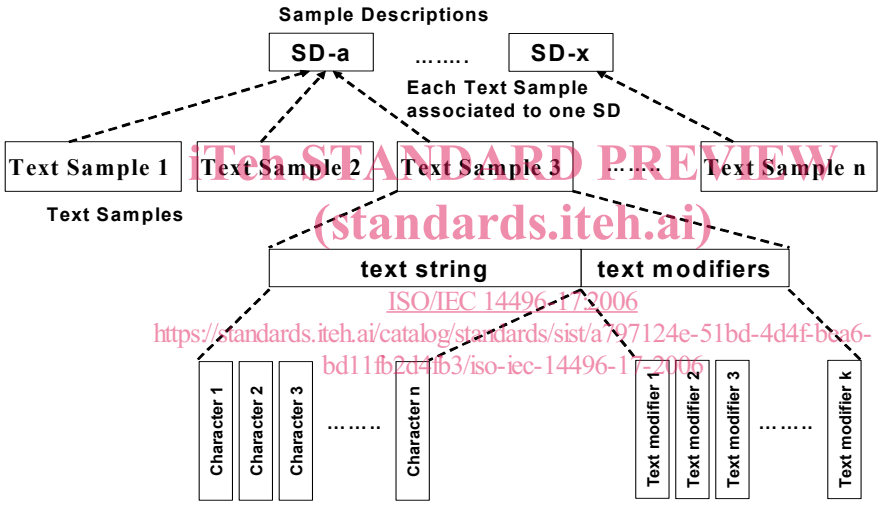

Figure 1 — Sample Descriptions, text samples, text strings and text modifiers in 3GPP text streams

A 3GPP text access unit contains data from exactly one text sample. Consequently, a 3GPP text access unit shall not contain data from more than one text sample. Each 3GPP text access unit is to be presented during a certain period of time, specified by duration information. In addition to a text sample, a 3GPP text access unit may contain zero or more complete sample descriptions. By allowing interleaving of text samples with sample descriptions in 3GPP text access units, 3GPP text streams are capable of carrying sample descriptions in-band[1)].

7.3 Transport of 3GPP text access units in TTUs

Typically, a 3GPP text access unit is small, around 100 – 200 bytes, and often much smaller than the size of the packets that carry the text data across a transport network. It is therefore expected that transport systems will often aggregate multiple 3GPP text access units into one transport packet. On the other hand, 3GPP text access units can also be large, for example when scrolling horizontal text at the bottom of the screen, in which case fragmentation of 3GPP text access units may be required prior to transport. In conclusion, transport of 3GPP text access units often requires aggregation and sometimes fragmentation. So as to conveniently

1) Note however that 3GPP sample descriptions may also be provided out-of-band.

aggregate and fragment 3GPP text access units in a transport independent manner, this Specification defines a flexible framing structure consisting of so-called TTUs, Timed Text Units.

Five different types of TTUs are defined; one for carriage of a complete 3GPP text access unit, three for carriage of text sample fragments, and one for carriage of a complete sample description, while three types are reserved for future use. Because sample descriptions are small, there is no support for carriage of sample description fragments.

The flexible framing structure provided by TTUs allows for easy and convenient adaptation to the various transport layers, while performing TTU alignment with the applied transport packets. For each transport layer the most suitable TTU structure can be chosen. For example, by using TTUs, small text samples can be aggregated into one transport packet, but TTUs can also be used to fragment text samples across multiple transport packets, while providing a reasonable level of error resilience in case of packet loss or non-recoverable packet errors. If so desired, the text data within a text access units can be re-partitioned into TTUs for most effective adaptation to other transport systems. See Figure 2. TTUs are defined for 3GPP text streams only, but comparable structures may be defined for other text streams in future versions of this Specification.

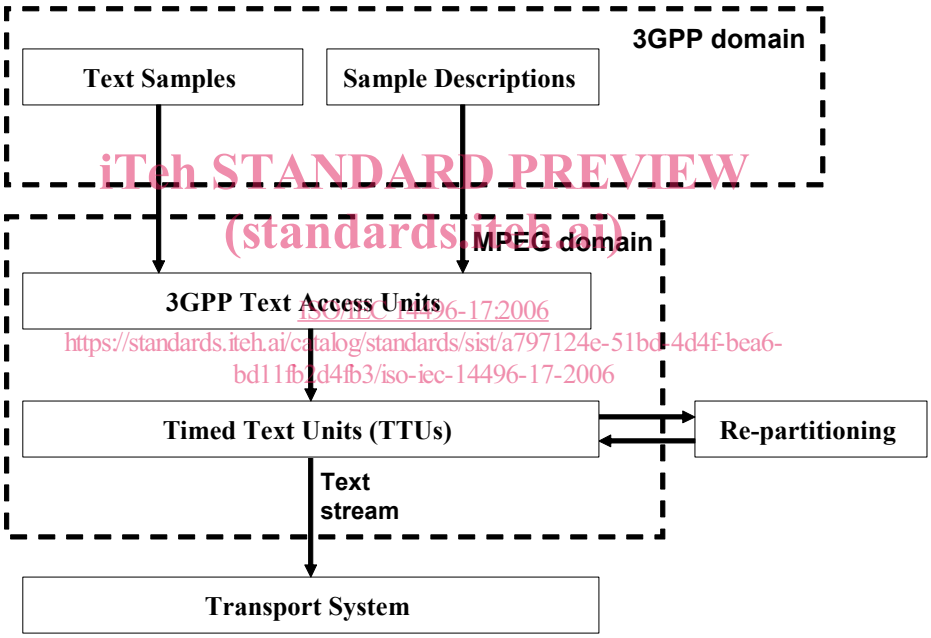

Figure 2 — Carriage of text samples and sample descriptions in 3GPP text access units and the use of TTUs for creating a 3GPP text stream

The format of TTUs is defined in this Specification, as well as some general requirements for their use, while the actual usage of TTUs is transport specific and beyond the scope of this Specification. However, as TTUs support text access unit aggregation and fragmentation in a generic and error resilient manner, transport systems are strongly recommended to use TTUs for aggregation and fragmentation instead of transport specific tools.

Each text access unit with a 3GPP text stream is defined to consist of one or more TTUs. Each TTU type is identified by its index j and referred to as TTU[j]. The following TTU types are defined:

- TTU[0]; this TTU type is reserved for future use;
- TTU[1]; this TTU type is capable of carrying one complete 3GPP text access unit;