

Australian Standard™

**Information technology—Coding of
audio-visual objects**

Part 17: Streaming text format



This Australian Standard was prepared by Committee IT-029, Coded Representation of Picture, Audio and Multimedia/Hypermedia Information. It was approved on behalf of the Council of Standards Australia on 19 April 2006.
This Standard was published on 8 May 2006.

The following are represented on Committee IT-029:

Australian Broadcasting Authority (ABA)
Australian Broadcasting Corporation (ABC)
Australian Consumers Association
Australian Subscription Television
CSIRO Mathematical & Information Services
Department of Defence (Australia)
Free TV Australia
School of Computer Science and Mathematics
Victoria University of Technology
Special Broadcasting Service (SBS)
The University of New South Wales
University of Wollongong

Keeping Standards up-to-date

Standards are living documents which reflect progress in science, technology and systems. To maintain their currency, all Standards are periodically reviewed, and new editions are published. Between editions, amendments may be issued. Standards may also be withdrawn. It is important that readers assure themselves they are using a current Standard, which should include any amendments which may have been published since the Standard was purchased.

Detailed information about Standards can be found by visiting the Standards Web Shop at www.standards.com.au and looking up the relevant Standard in the on-line catalogue.

Alternatively, the printed Catalogue provides information current at 1 January each year, and the monthly magazine, *The Global Standard*, has a full listing of revisions and amendments published each month.

Australian Standards™ and other products and services developed by Standards Australia are published and distributed under contract by SAI Global, which operates the Standards Web Shop.

We also welcome suggestions for improvement in our Standards, and especially encourage readers to notify us immediately of any apparent inaccuracies or ambiguities. Contact us via email at mail@standards.org.au, or write to the Chief Executive, Standards Australia, GPO Box 476, Sydney, NSW 2001.

This Standard was issued in draft form for comment as DR 06075.

Australian Standard™

Information technology—Coding of audio-visual objects

Part 17: Streaming text format

First published as AS ISO/IEC 14496.17—2006.

COPYRIGHT

© Standards Australia

All rights are reserved. No part of this work may be reproduced or copied in any form or by any means, electronic or mechanical, including photocopying, without the written permission of the publisher.

Published by Standards Australia GPO Box 476, Sydney, NSW 2001, Australia

ISBN 0 7337 7398 2

PREFACE

This Standard was prepared by the Standards Australia Committee IT-029, Coded Representation of Picture, Audio and Multimedia/Hypermedia Information.

This Standard is identical with, and has been reproduced from ISO/IEC 14496-17:2006, *Information technology—Coding of audio-visual objects, Part 17: Streaming text format*.

The objective of this Standard is to provide the Australian multimedia industry with a generic method for coding of text at very low bitrate, for creation of subtitles and karaoke song texts to be coded and transported as separate text streams for presentation jointly with other components of an audiovisual presentation for use in mobile services over IP.

Standards Australia is in the process of adopting most of ISO/IEC 14496 series Standards. Refer to the website for current list.

The terms ‘normative’ and ‘informative’ are used to define the application of the annex to which they apply. A normative annex is an integral part of a standard, whereas an informative annex is only for information and guidance.

As this Standard is reproduced from an international standard, the following applies:

- (a) Its number appears on the cover and title page while the international standard number appears only on the cover.
- (b) In the source text ‘this International Standard’ should read ‘this Australian Standard’.
- (c) A full point substitutes for a comma when referring to a decimal marker.

References to International Standards should be replaced by references to Australian or Australian/New Zealand Standards, as follows:

<i>Reference to International Standard</i>		<i>Australian Standard</i>	
ISO/IEC		AS ISO/IEC	
14496	Information technology – Coding of audio-visual objects	14496	Information technology – Coding of audio-visual objects
14496-18	Part 18: Font compression and streaming	14496.18	Part 18: Font compression and streaming

Only international references that have been adopted as Australian or Australian/New Zealand Standards have been listed.

CONTENTS

Page

1	Scope	1
2	Normative references	1
3	Terms and definitions.....	1
4	Text stream format.....	2
5	Text access units	2
5.1	Timing and decoding of text access units	2
5.2	Format of text access units	2
5.3	TextConfig	3
5.4	Semantics	3
6	Usage of a text stream within an MPEG-4 system context.....	3
6.1	Signaling of a text stream	3
6.2	Usage in the scene description.....	3
7	Text data format for 3GPP text streams	3
7.1	Introduction	3
7.2	Carriage of text samples and sample descriptions in 3GPP text access units.....	4
7.3	Transport of 3GPP text access units in TTUs.....	4
7.4	TTU Syntax and Semantics.....	9
7.5	Positioning of 3GPP text streams	13
7.6	Format specific decoder configuration for 3GPP text streams	14
7.7	Hypothetical Text Decoder for 3GPP text stream decoding.....	17
7.8	Profile and Level parameters for 3GPP text streams	18
	Annex A (normative) Font Referencing	19
	Annex B (informative) Transport of Text Streams	20
	Bibliography	21

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 sufficiently low for use in mobile services over IP.

AUSTRALIAN STANDARD

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.

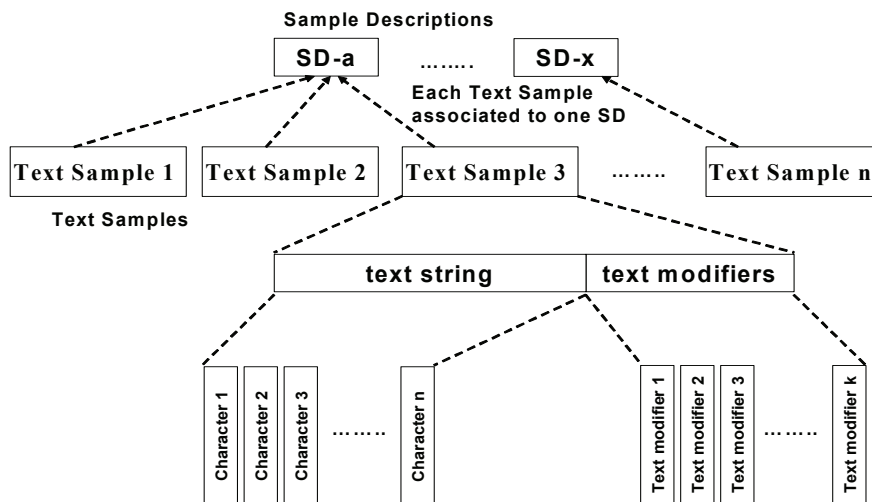


Figure 1 — Sample Descriptions, text samples, text strings and text modifiers in 3GPP text streams

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-band¹⁾.

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.