

SMPTE STANDARD

Timed Text Format
(SMPTE-TT)



Table of Contents	Page
Foreword.....	2
Intellectual Property.....	2
Introduction	2
1 Scope.....	3
2 Conformance Notation.....	3
3 Normative References	3
4 Definitions and Acronyms.....	4
4.1 Definitions	4
4.2 Acronyms.....	4
5 SMPTE TT-Format	5
5.1 Translation Modes	5
5.2 Default Styles.....	6
5.3 SMPTE Namespaces	6
5.4 Tunneling Transport	6
5.5 Pre-rendered Backgrounds	7
5.6 Font Resolution	11
5.7 SMPTE Metadata XML Vocabulary.....	11
5.8 SMPTE-TT Features from TTML.....	14
5.9 SMPTE-TT Specific Features.....	14
Annex A Bibliography (Informative).....	15
Annex B Chunking (Informative).....	16
Annex C XSD Module of SMPTE-TT Extensions (Informative)	18

Foreword

SMPTE (the Society of Motion Picture and Television Engineers) is an internationally-recognized standards developing organization. Headquartered and incorporated in the United States of America, SMPTE has members in over 80 countries on six continents. SMPTE's Engineering Documents, including Standards, Recommended Practices, and Engineering Guidelines, are prepared by SMPTE's Technology Committees. Participation in these Committees is open to all with a bona fide interest in their work. SMPTE cooperates closely with other standards-developing organizations, including ISO, IEC and ITU.

SMPTE Engineering Documents are drafted in accordance with the rules given in Part XIII of its Administrative Practices.

SMPTE ST 2052-1 was prepared by Technology Committee 24TB.

Intellectual Property

At the time of publication no notice had been received by SMPTE claiming patent rights essential to the implementation of this Engineering Document. However, attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. SMPTE shall not be held responsible for identifying any or all such patent rights.

Introduction

This section is entirely informative and does not form an integral part of this Engineering Document.

Timed text is an important essence element of information and entertainment programming. Various regulatory bodies and administrations around the world have, in both technical specification and quantitative requirement, made various forms of timed text (such as captions or subtitles) a necessary component of content.

This Standard provides a framework for timed text to be supported for content delivered via broadband means, taking into consideration the economic and practical necessity of the ability to use source data delivered in pre-existing and regionally-specific formats (such as CEA-708, CEA-608, DVB Subtitles, and WST (World System Teletext)) as well as text that may be authored or provided in other formats.

It is a goal of this Standard to make timed text essence more useful for broadband content, with equal status to the associated video and audio essence. Considering however that contemporary broadband content on many platforms (e.g., media players, media extenders, home servers, home networks, etc.) may be intended for display in a consumer electronic environment, and that there will be a transition period where such equipment can only utilize pre-existing formats, this Standard also defines functionality for the preservation of legacy formats. It is the intent, however, that new display devices will utilize the more capable TTML timed text structures rather than legacy data formats.

1 Scope

This Standard defines the SMPTE profile of W3C Timed Text Markup Language (TTML), designated SMPTE-TT, which may represent Captions or Subtitles. It identifies the features from TTML required for interoperability between display systems for the format. In order to preserve certain semantic features of the input formats, SMPTE-TT also defines some standard metadata terms to be used, and some extension features not found in TTML.

2 Conformance Notation

Normative text is text that describes elements of the design that are indispensable or contains the conformance language keywords: "shall", "should", or "may". Informative text is text that is potentially helpful to the user, but not indispensable, and can be removed, changed, or added editorially without affecting interoperability. Informative text does not contain any conformance keywords.

All text in this document is, by default, normative, except: the Introduction, any section explicitly labeled as "Informative" or individual paragraphs that start with "Note:"

The keywords "shall" and "shall not" indicate requirements strictly to be followed in order to conform to the document and from which no deviation is permitted.

The keywords, "should" and "should not" indicate that, among several possibilities, one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required; or that (in the negative form) a certain possibility or course of action is deprecated but not prohibited.

The keywords "may" and "need not" indicate courses of action permissible within the limits of the document.

The keyword "reserved" indicates a provision that is not defined at this time, shall not be used, and may be defined in the future. The keyword "forbidden" indicates "reserved" and in addition indicates that the provision will never be defined in the future.

A conformant implementation according to this document is one that includes all mandatory provisions ("shall") and, if implemented, all recommended provisions ("should") as described. A conformant implementation need not implement optional provisions ("may") and need not implement them as described.

Unless otherwise specified, the order of precedence of the types of normative information in this document shall be as follows: Normative prose shall be the authoritative definition; Tables shall be next; followed by formal languages; then figures; and then any other language forms.

3 Normative References

The following standards contain provisions which, through reference in this text, constitute provisions of this Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this Standard are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

IETF RFC 2396 (August 1998), Uniform Resource Identifiers (URI): Generic Syntax

IETF RFC 4648 (October 2006), The Base16, Base32, and Base64 Data

ISO/IEC 15948:2003. Information Technology — Computer Graphics and Image Processing — Portable Network Graphics (PNG): Functional Specification

W3C Timed Text Markup Language (TTML) 1.0, W3C Recommendation 18 November 2010
<http://www.w3.org/TR/ttaf1-dfxp/>

W3C Extensible Markup Language (XML) 1.0 (Fifth Edition), W3C Recommendation 26 November 2008

W3C Extensible Stylesheet Language (XSL) Version 1.1, W3C Recommendation 05 December 2006

4 Definitions and Acronyms

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

4.1 Definitions

Caption: textual representation of the audio track, usually including all sounds, and usually in the same language as the audio track dialog, intended for hearing impaired audiences.

Chunk: A segment of a media timeline that forms part of a larger contiguous sequence of timeline segments where each segment has a 1 to 1 mapping with a SMPTE-TT document.

Closed Captions: Captions that may be displayed or not displayed depending on user preference.

Datatype: A metadata description defining the encoding format of data.

Div: A container section of a TTML document as defined by the W3C TTML standard.

Infoset: an abstract Data Model describing the information available from an XML document.

Namespace: mechanism for scoping the definitions of names in an XML document.

Presentation processor: Hardware or software that creates a visible rendition of a document as defined in W3C TTML.

Subtitle: textual representation of the audio track, usually just the dialog and usually in a language other than the audio track dialog, intended for foreign language audiences.

SMPTE-TT document: an encoding of some XML content conforming to the SMPTE-TT document type.

Timed Text: generically all text data sources, including Captions and Subtitles.

XInclude: A mechanism for including some XML content within another XML document.

4.2 Acronyms

CEA: Consumer Electronics Association

DVB: Digital Video Broadcasting

MIME : Multipurpose Internet Mail Extensions

PNG: Portable Network Graphics

SMPTE-TT: SMPTE Profile of TTML

TTML (W3C): Timed Text Markup Language

URI: Universal Resource Identifier

W3C: World Wide Web Consortium

XML: eXtensible Markup Language

XSL: Extensible Stylesheet Language

5 SMPTE-TT Format

5.1 Translation Modes

5.1.1 Requirements

A SMPTE-TT document describes closed caption or subtitle data for association with a given media asset. Such documents can either be created directly or translated from an existing data source. In this Standard, the term SMPTE-TT document is used to describe either a directly created document or one translated from an existing data source, which will be referred to in this Standard as a legacy format.

In order for users to correctly interpret a translation of a legacy format, it may be required that all display features of the original are preserved in case they are being used to convey meaning. Similarly, the timing of data extraction and preservation of latency, display modes, and any other features that affect a display “instant” may also be crucial. Translation accuracy may be required to ensure that no meaning is missed by the viewer after the translation.

SMPTE-TT, however, allows for features, especially in the area of internationalization and styling that cannot be specified in some legacy formats such as CEA-608. In some cases, it may be preferable to replace legacy features with new presentation modes where this does not affect the meaning of the content.

In order that additional features of SMPTE-TT may be used when translating, this Standard defines two modes of translation: Preserved and Enhanced to meet these two cases.

5.1.2 Modes

The mode used for translation shall be recorded in the `smpte:information` element as defined in Section 5.7.4. Each translated document shall use one of these modes exclusively. Specifics of what constitutes each of these modes for a given legacy format is outside the scope of this Standard and is addressed in the RPs defining the mappings from other formats into SMPTE-TT.

5.1.2.1 Preserved mode

In Preserved mode, all legacy data shall be correctly converted (e.g., CEA-608 pop-on, roll-up, etc.) such that the resulting TTML rendering is as accurate a representation of the original rendering as possible, given the constraints of the formats.

In Preserved mode, if some presentation feature of the original file cannot be captured by any SMPTE-TT feature, this shall be indicated in the header of the file using the `smpte:information` element as defined in Section 5.7.4. The translation shall continue to preserve as much as possible of the remainder of the document.

5.1.2.2 Enhanced mode

In Enhanced mode, SMPTE-TT shall be used to convey the semantic information of the legacy data, but may differ in its exact presentation. Legacy data may be more liberally interpreted so that exact presentation

fidelity may not be preserved (e.g., matching of visual effects, typography, and coloring may not be required). All content presented in the same style in the legacy data however, shall be translated to a common destination style.

In Enhanced mode, exact output positioning need not be preserved, but may be mapped to approximately equivalent grouped areas (e.g., upper third, lower third).

In Enhanced mode, exact timing need not be preserved (e.g., in the case of CEA-608, frame accurate replication of onset of each pair of letters is not required, and all characters in a line may be aggregated into a single onset time where the onset time of the aggregated group is equal to the original onset time of first visible character in the group, and the outgoing time of the aggregated group, the outgoing time of the last character in the group).

5.2 Default Styles

TTML leaves some initial style values to be defined in context. When computing style inheritance, the set of initial style values shall be set according to Table 1.

Table 1 – Initial style values

tts:color	white
-----------	-------

When a SMPTE-TT document with no region defined (using the TTML default region) is presented, the style initial values of Table 2 shall be used.

Table 2 – Initial style values – no region defined

tts:displayAlign	after
tts:textAlign	center

5.3 SMPTE Namespaces

TTML allows for metadata to be added in private namespaces. SMPTE-TT documents need not use the prefixes in Table 3 although such prefixes are used in examples throughout this specification. It is assumed that if a full namespace declaration is not given in an example XML fragment then the definitions in Table 3 are set in an outer context.

Table 3 – Namespaces

Name	Prefix	Value
SMPTE	smpte:	http://www.smpte-ra.org/schemas/2052-1/2013/smpte-tt