INTERNATIONAL STANDARD

ISO/IEC 14496-30

Second edition
2018-11

AMENDMENT 1
2022-06

Information technology — Coding of audio-visual objects —

Part 30: Timed text and other visual overlays in ISO base media file format

AMENDMENT 1: Timing improvements

Technologies de l'information — Codage des objets audiovisuels —

Partie 30: Texte temporisé et autres recouvrements visuels dans le format ISO de base pour les fichiers médias

AMENDEMENT 1: Améliorations des temporisations

Reference number
ISO/IEC 14496-30:2018/Amd. 1:2022(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Information technology — Coding of audio-visual objects —

Part 30: Timed text and other visual overlays in ISO base media file format

AMENDMENT 1: Timing improvements

Clause 3

Replace the 3.1.2 entry with

timed text stream

stream of content, which when decoded results in textual content, possibly containing internal timing values, to be processed at a given presentation time and for a certain duration

4.2, second paragraph

Replace the second sentence of the second paragraph, "The rendering of the sample happens at the composition time, taking into account edit lists if any, and for the whole sample duration, without timing behaviour.", with:

The rendering of the sample happens at the presentation time, i.e. taking into account edit lists if any, and for the sample duration, potentially trimmed by the edit list if any.

NOTE As defined in ISOBMFF, the presentation is driven by the presentation time. The composition time is determined from the presentation time using the edit list (if present), and the sample active at that composition time is then processed. This specification assumes that at a given presentation time, the renderer is provided with the sample and the composition time, both of which correspond to the given presentation time.

5.3

Replace the entire subclause with:

This subclause defines processing of timing information for TTML documents carried in a TTML track. The general timing processing defined in 4.2 applies, but specific aspects are refined in this subclause. Timing processing is defined only for TTML documents in which ttp:timeBase is 'media'. For other values, timing processing behaviour is undefined and such documents should not be carried in TTML tracks.

When the rendering of a sample happens, the TTML document in the sample is provided to the TTML processor, together with the track composition time T of the sample and the sample composition duration d. The TTML processor then uses the interval $[T, T+d)$ together with the time coordinates $\{T_i\}$ produced by the "resolve timing" procedure, as defined in TTML, to determine which intermediate

synchronic documents (ISDs) to present and for how long. Specifically, track composition time $t \in [T,\ T+d)$ results in TTML ISD i being presented, with $t \in [T_i,\ T_{i+1})$.

NOTE 1 The fact that time coordinates produced by the TTML processor are interpreted as being on the track composition timeline remains true in the case of segment files, defined in 14496-12. In other words, time coordinates produced by a document stored in a sample of a segment are still relative to time 0 on the track composition timeline and are not relative to the segment start.

The above concepts are illustrated in Table 1 and Table 2.

Table 1 — Example of a TTML track with six samples

Sample	Composition time	Duration	Payload
1	00:00:00	00:30:00	<tt> <body> 1-2 minutes </body> </tt>
2	00:30:00	00:30:00	<tt> <body> 31-32 minutes </body> </tt>
3	01:00:00	00:30:00	<tt> <body> 60-150 minutes </body> </tt>
4	01:30:00	00:30:00	<tt> <body> 60-150 minutes </body> </tt>
5	02:00:00	00:30:00	<tt> <body> 60-150 minutes </body> </tt>

Table 1 *(continued)*

Sample	Composition time	Duration	Payload
6	02:30:00	00:30:00	<tt> <body> 150 minutes onwards </body> </tt>

The sample composition times of the samples in Table 1 are 0 min, 30 min, 1 h, 1 h 30 min, 2 h, and 2 h 30 min, which correspond to the time at which the decoder processes the TTML content from that sample. The text content in the payload of these samples reflects when that text will be displayed. For example, the text "60-150 minutes" is expected to be shown between composition times 60 min and 150 min. The timing information in the TTML documents in these samples is not necessarily matching the times in the text content (e.g. 00:00:00 to 04:00:00) precisely to illustrate the impact of storing the content in ISOBMFF samples. Table 2 shows how each sample is processed to produce ISDs, and then how these are clipped to the sample times for display. The ISDs excerpts in Table 2 are illustrative and only focusing on timing aspects and content. Compliant TTML ISDs can need to include region elements.

Table 2 — ISD for the TTML track of Table 1

Sample	ISD contents produced by TTML's "resolve timing" procedure and associated time range	ISD contents output from sample processing and associated time range
1		00:00:00 to 00:01:00 Empty document
	00:01:00 to 00:02:00 1-2 minutes	00:01:00 to 00:02:00 1-2 minutes
	00:02:00 to Infinity Empty document	00:02:00 to 00:30:00 Empty document
2		00:30:00 to 00:31:00 Empty document
	00:31:00 to 00:32:00 31-32 minutes	00:31:00 to 00:32:00 31-32 minutes
	00:32:00 to Infinity Empty document	00:32:00 to 01:00:00 Empty document
3	00:00:00 to 04:00:00 60-150 minutes	01:00:00 to 01:30:00 60-150 minutes
	04:00:00 to Infinity Empty document	
4	00:00:00 to 04:00:00 60-150 minutes	01:30:00 to 02:00:00 60-150 minutes
	04:00:00 to Infinity Empty document	
5	02:00:00 to 02:30:00 60-150 minutes	02:00:00 to 02:30:00 60-150 minutes
	02:30:00 to Infinity Empty document	

Table 2 *(continued)*

Sample	ISD contents produced by TTML's "resolve timing" procedure and associated time range	ISD contents output from sample processing and associated time range
6	02:30:00 to Infinity 150 minutes onwards	02:30:00 to 03:00:00 150 minutes onwards

When processing sample 1, following TTML's "resolve timing" procedure, two ISDs are created: one non-empty ISD ISD_1 and one subsequent empty ISD ISD_2. However, because the time interval of ISD_1 starts after the sample timing interval, the sample processor can additionally output one more ISD, corresponding to an empty document, to be displayed from the start of the track until the start of ISD_1. ISD_2 is to be displayed from the end of ISD_1 until the sample composition end time is reached. The sample processor can omit creating these empty-document ISDs and clear the display when there is no ISD for the given time ranges.

Sample 2 is processed similarly to Sample 1. However, since it is provided to the decoder at composition time 30 min and since the sample composition duration is 30 min, the TTML decoder will display the ISD corresponding to the first empty document only during the interval [00:30:00, 00:31:00), the non-empty intermediate synchronic document during the interval [00:31:00, 00:32:00) and finally display the ISD corresponding to the last empty document until the sample composition time is reached.

Samples 3 and 4 contain exactly the same content and illustrate the possibility of having duplicate content in different samples. This can happen for example at the boundaries of segments in segmented files. Both samples will produce the same ISD. In order to avoid possible rendering artefacts such as flickering at the boundary between such samples, decoder implementations can detect such duplicate, adjacent ISDs and keep processing the first ISD. Moreover, for these samples, TTML's "resolve timing" procedure will produce two ISDs: one with a time interval that is larger than the interval of their containing sample, and one corresponding to an empty document with a time interval outside of the sample interval. As indicated in Table 2, the sample processor will ignore the second ISDs and clip the rendering of each document in the first ISDs to the sample interval.

Sample 5 also contains the same text content as Sample 3 and 4, but the timing values in the document have been adjusted to match the containing sample interval. Both types of samples result in identical output in this case and both are valid. Sample 5 follows a recommendation that was present in the previous version of this specification that the duration of a TTML document carried in a sample should not be greater than the sample duration.

Finally, Sample 6 illustrates the case of a document that has no end time. The TTML processor will produce one non-empty ISD with time interval [02:30:00, Infinity). However, the sample processor will display it for the clipped interval defined by the sample composition time and duration.

5.6

At the end of the subclause, add the following text:

If a sample contains the identical document to the prior sample, it can be marked as redundant, with the `sample_has_redundancy` flag defined in ISO/IEC 14496-12, allowing processors the option to extend the duration of the prior sample before processing it and then discard the new sample.

An 'empty' sample is defined as containing a TTML document that has no content. A TTML document that has no content is any document that contains (a) no `<div>` element or (b) no `<body>` element or (c) no `<p>` or `<span>` elements containing character data or `<br>` elements; for example, the following document:

```
<tt xml:lang="" xmlns="http://www.w3.org/ns/ttml"/>
```

5.9

Replace this entire subclause with:

a) The earliest begin time of an element in a TTML document can be non-coincident (earlier or later) with the composition time of the containing sample, and the latest computed end time of an element in the TTML document can be non-coincident (earlier or later) with the end of the sample, which is equal to its composition time plus sample duration.

b) TTML content that falls partially or wholly within the duration of a sample can be present (duplicated) in adjacent samples.

c) Only one sample and document within a Timed Text Track can be active at any moment in the presentation. The presentation of every document is constrained in time to the period beginning at the composition time of the containing sample with the duration of that sample. As a consequence, any timed content within a document that extends outside of that period is expected to be temporally clipped and not to result in any impact on presentation prior to or later than the sample's composition time and duration.

5.10

Create a subclause 5.10 Profiles and buffering constraints with the following existing text:

The present document does not define a transport layer buffer or timing model. Such a model can be used to guarantee that subtitle content can be read and processed in time to be synchronously presented with audio and video. It is assumed that users of this track format will define timed text content profiles and hypothetical render models (HRM) that will constrain content parameters so that compatible decoders can identify and decode those profiles for synchronous presentation. For example, the IMSC1 profiles define such an HRM.

When a track carries IMSC1 documents, and for the purpose of verifying HRM constraints, the sequence of ISDs to be used by the algorithm defined in IMSC1 is the one made of all the ISDs produced by all the sample processing defined in 5.3, i.e. after potential clipping by the sample interval.

The following non-exhaustive list of document constraints may need to be specified to define a timed text profile that will guarantee synchronous decoding of conforming content on conforming decoders:

— maximum allowed document size;

— number of document buffers in the hypothetical render model;

— video overlay timing of the hypothetical render model;

— maximum total compressed image size in megabytes per sample;

— maximum total decoded image size in megapixels per sample;

— maximum decoded image dimensions;

— maximum text rendering rate required by a document;

— maximum image rendering rate required by a document;

— maximum number of simultaneously displayed characters;

— maximum font size;

— maximum number of simultaneously displayed images.

NOTE Defining timed text content profiles is outside the scope of this document but providing a method to signal an externally defined timed text profile in the subtitle sample description is possible using the sample entry description.

6.3

At the end of this subclause, add the following text:

When a file writer identifies a cue with indefinite duration, since the cue has no end-time, setting an accurate value for the sample duration might not be possible. In that case, the file writer may rely on tools defined in ISO/IEC 14496-12. If the sample is the last one in the track, ISO/IEC 14496-12:2015, 8.6.1.1 permits setting a zero duration for that sample. If the sample is not the last one in the track, ISO/IEC 14496-12 defines cases when a sample duration can be extended by subsequent information: either using redundant samples (see ISO/IEC 14496-12:2015, 8.6.4.1); or in fragmented tracks, if the sample is the last in the fragment using the `TrackFragmentDecodeTime` box of the next fragment (see ISO/IEC 14496-12:2015, 8.8.12.1). If such sample duration extension is used, setting an accurate duration is not needed and the file writer may choose an arbitrary short duration for the sample.

Additionally, if the sample containing a cue with indefinite duration is not the last one, since every sample in a WebVTT track is a sync sample, file writers shall duplicate the indefinite cue content in all subsequent samples.

6.5, NOTE

Change the syntax of the class WVTTSampleEntry to:

```
class WVTTSampleEntry() extends PlainTextSampleEntry ('wvtt'){
   WebVTTConfigurationBox   config;
   WebVTTSourceLabelBox     label;   // recommended
   BitRateBox               bitrate;
                               // optional, as defined in ISO/IEC 14496-12
}
```

Bibliography

Add the following reference at the end of the Bibliography:

[6] IMSC1, TTML Profiles for Internet Media Subtitles and Captions, https://www.w3.org/TR/ttml-imsc/rec