Subtitle Safe Crop Area SCA

BBC, 9th June 2016

Introduction

This document describes a proposal for a Safe Crop Area parameter attribute for inclusion within TTML documents to provide additional information to presentation processors to improve the options available when using TTML for subtitles or captions positioned alongside a related video object in the case where the related video object is not fully visible, for example because the viewer's preference is to fill the display even if that means part of the related video frame is no longer visible.

Throughout this document the name "**Safe Crop Area**" is used however other names can be considered – one alternative that has been proposed is "**Subtitle Editorial Area**".

Problem statement

When using TTML for subtitles and captions of video, it is important to preserve the relative positioning of the text against the video. Features such as IMSC's `ittp:aspectRatio` or the EBU-TT rule that the reference for the origin and extent of regions is the active video define this spatial relationship. (Note that the EBU-TT-D rule is less tightly defined)

However there are many instances where display systems such as televisions, tablets, HTML5 video elements etc. do not show all of the active video. There are many reasons why this occurs. Televisions traditionally overscan the image slightly, cutting off the edges. Many displays do not have the same aspect ratio as the active video, and permit the user to choose whether to display all of the video with black bars top and bottom or left and right, or to zoom the video, cropping some edges.

It is also important to ensure that, where subtitles and captions are important, the text is displayed. If a rendered TTML region extends into an area of the active video that is being cropped, some words may not be readable or displayed at all.

In the broadcast video world some hints are provided using Active Format Descriptors (AFDs) to help presentation devices to make good choices, for instance by telling the display when the video is in fact 4:3 aspect ratio within a 16:9 aspect ratio frame, so that the appropriate display on a 4:3 display is to centre-cut the video.

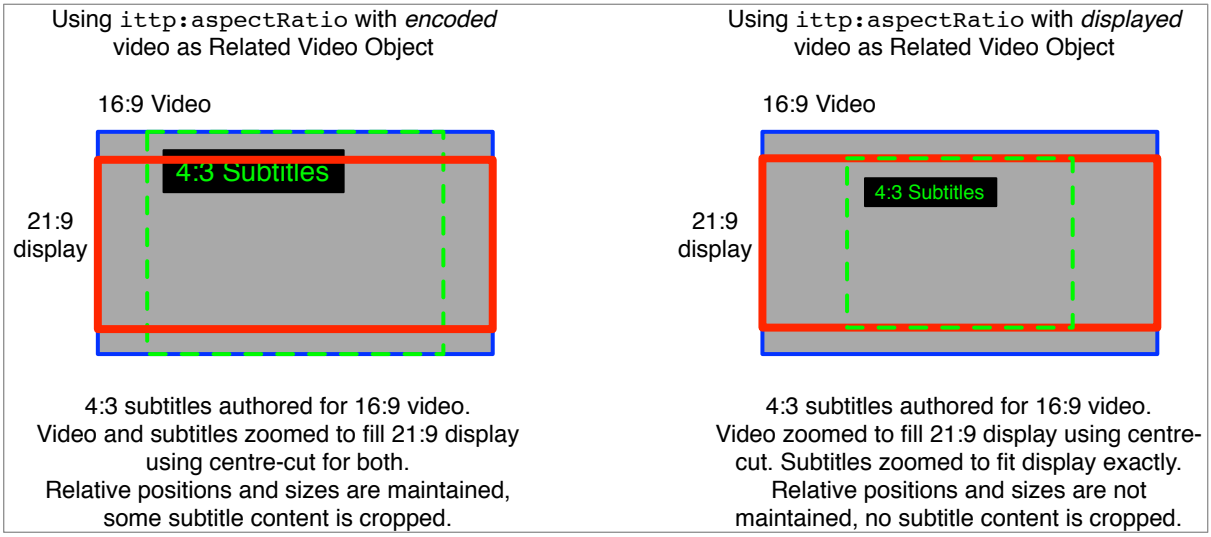
There is currently no such equivalent for TTML, so a presentation processor has some choices, none of which is ideal:

- Maintain alignment between the root container region and the active video according to the specification, and risk clipping some text;
- Scale the root container region to the display size and lose positional accuracy, and risk introducing line breaks and/or overflow scenarios;
- Perform complex analysis of the TTML content to determine how to display the content overriding the positional information within the document instance.

It has been suggested that using `ittp:aspectRatio` could be sufficient to solve this problem: the document can be authored within, say, a 4:3 aspect ratio root container region that is centred relative to the display. This does indeed work in a limited set of scenarios. However that approach fails if the display is not showing the entire video frame that coincides with root container.

For example, consider a 16:9 video for which 4:3 subtitles are authored using `ittp:aspectRatio="4 3"`, which is then displayed on a 21:9 display which the user has configured to fill, cropping the top

and bottom of the video. If the Related Video Object frame is taken to be the encoded video, then the top and bottom of the root container region will also be cropped; conversely if the Related Video Object frame is taken to be the displayed part of the video, then the relative positions and sizes of the TTML content will not as authored. Neither behaviour is ideal. See the following diagram:



The desired semantics are:

a) to consistently define the relationship between the root container region and the encoded video frame, so that relative positions and sizes as authored can be maintained;

b) to signal that part of the root container region which must not be cropped.

This combination would allow document authors to target a range of display aspect ratios regardless of whether all of the video is being shown or some cropping is occurring. It would also allow presentation systems to make better decisions about whether and how much to crop the root container region or apply some translation and scaling, or some combination of the two.

Note: we do not present in this document any requirements for an algorithm for such translation and scaling; the assumption is that the entire safe crop area must be displayed, maintaining as much spatial alignment with video as is possible.

References

[IMSC1] TTML Text and Image Profiles for Internet Media Subtitles and Captions 1.0: https://www.w3.org/TR/ttml-imsc1/

[TTML1] TTML 1 Second Edition: https://www.w3.org/TR/ttml1

[TTML2] TTML 2 Editor's Draft: https://rawgit.com/w3c/ttml2/master/spec/ttml2.html

Definitions

The following terms are used in this proposal:

Interchange Context — Synonym for Document Interchange Context as defined in [TTML1]:

> The implied context or environment external to a *Content Processor* in which document interchange occurs, and in which out-of-band protocols or specifications may define certain behavioral defaults, such as an implied profile.

Presentation Processor — As defined in [TTML1]:

> A *Content Processor* which purpose is to layout, format, and render, i.e., to present, *Timed Text Markup Language* content by applying the presentation semantics defined in this specification.

Processing Context — Synonym for Document Processing Context as defined in [TTML1]:

> The implied context or environment internal to a *Content Processor* in which document processing occurs, and in which out-of-band protocols or specifications may define certain behavioral defaults, such as the establishment or creation of a *Synthetic Document Syncbase*.

Region — As defined in [TTML1]:

> A logical construct that models authorial intention regarding desired or potential presentation processing, and which is represented as a rectangular area of a presentation surface into which content is composed and rendered during presentation processing.

Related Media Object — As defined in [TTML1]:

> A (possibly null) media object associated with or otherwise related to a Document Instance. For example, an aggregate audio/video media object for which a Document Instance provides caption or subtitle information, and with which that Document Instance is associated.

Related Media Object Region — As defined in [TTML1]:

> When a non-null Related Media Object exists, the region of this media object, expressed in the coordinate system that applies to the Document Instance that is associated with the related media object.

Related Video Object — As defined in [IMSC1] §6.5:

> A Document Instance MAY be associated with a related video object, which SHALL consist of a sequence of image frames, each a rectangular array of pixels, and SHALL be considered the Related Media Object.

Root Container Region — As defined in [TTML1]:

> A logical region that establishes a coordinate system into which *Document Instance* content regions are placed and optionally clipped.

Safe Crop Area — Defined in this document:

> The rectangular portion of the Root Container Region identified as being unsuitable for cropping by the Presentation Processor.

Visible Rendering Region — Defined in this document:

> A Related Media Object Region for the Related Media Object that is the rectangular area of the visual display medium within which the TTML content is being presented; this would normally be coincident with the area within which any Related Video Object is being presented; it can but does not necessarily fully enclose the entire Root Container Region.
>
> Note: scaling or translating the presentation can change the Visible Rendering Region.
>
> Note: EBU-TT-D requires that implementations provide a "rendering plane" that is coincident with the root container region. The Visible Rendering Region differs from this since it can be either larger or smaller than the root container region.

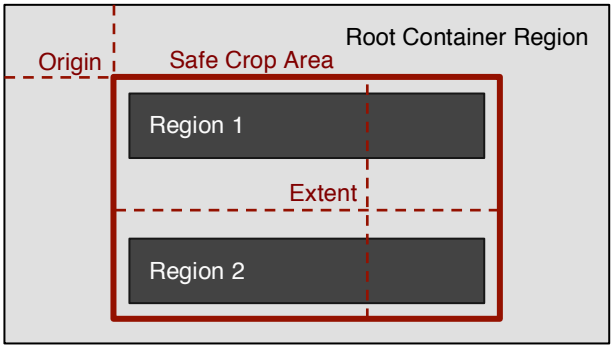
Note: it is possible for the Visible Rendering Region to extend outside the root container region, e.g. to have the equivalent of a negative origin or a size greater than 100% in one or both dimensions, however this does not imply an expectation that content which overflows the root container region will not be clipped to the limits of the root container region.

Conformance requirements on TTML documents

The attribute `[namespace]:safeCropArea` indicates the area that a presentation processor should display without cropping.

The `safeCropArea` attribute allows the content author to define the portion of the root container region that needs to remain visible. Typically the Safe Crop Area would be an area that fully contains all of the referenced regions within the content.

The Safe Crop Area is defined in the coordinate space of the document instance, and has an origin and extent. If the origin is omitted then the Safe Crop Area shall be located in a centred position relative to the root container region.



If omitted, the Safe Crop Area shall be coincident with the root container region.

The Safe Crop Area should fully enclose all Regions that are referenced within the content in the document instance.

Note: this allows content providers to supply documents that contain content outside the safe crop area if they determine that cropping such content is acceptable. An example where this may arise would be a presentation style that places all text in a full width region with a visible background colour. If the text is known to occupy only a portion of that region then only that portion where text will be presented needs to be in the safe crop area, since cropping part of the background leads to no loss of information.

In cases where the content for a related video is supplied in a sequence of TTML document instances the Safe Crop Area should not vary between document instances where the related video is editorially similar and should fully enclose all Regions that are referenced within the content in the whole sequence of document instances. For example if multiple documents contain captions or subtitles for a single programme whose aspect ratio (including black bars etc.) do not change then the Safe Crop Area should remain consistent across all of those documents. This avoids potentially jarring receiver behaviour that dynamically resizes text across document boundaries.

Note: the related video for a sequence of TTML documents depends on the context; typically it could be considered a single programme, rather than the entire continuous stream of a live rolling new channel for example.

Note: any regions that fall wholly or partially outside the safe crop area could be clipped by the presentation processor.

Syntax

The `[namespace]:safeCropArea` attribute is defined as an optional attribute of the `<tt>` element.

If present, the `[namespace]:safeCropArea` attribute shall conform to the following syntax:

```
[namespace]:safeCropArea
  : <origin> <whitespace> <extent>

<origin>
  : <x> <whitespace> <y>

<extent>
  : <width> <whitespace> <height>

<width> | <height> | <x> | <y>
  : <length>

<whitespace>
  : (#x20 | #x9 | #xD | #xA)+
```

where `<length>` expressions are as defined in [TTML1] §8.3.9.

Width and x percentage values shall be relative to the width of the root container region.

Height and y percentage values shall be relative to the height of the root container region.

The safe crop area shall not extend outside the root container region in any dimension.

Presentation Processor behaviour

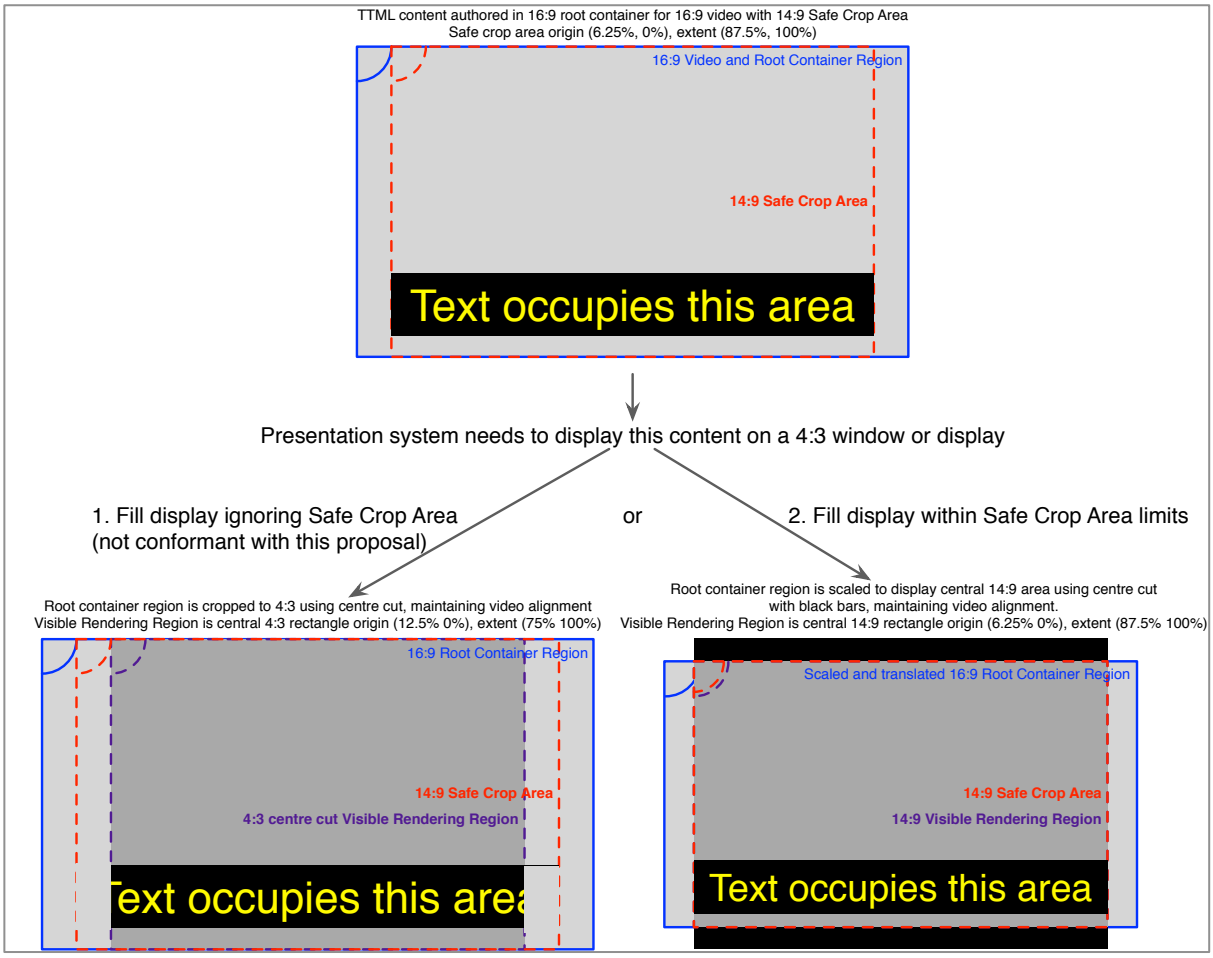
Presentation Processors shall display the TTML content that falls within the safe crop area. Presentation Processors should use best endeavours to maintain text size and positional alignment of TTML content within the safe crop area relative to the Related Video Object.

As a minimum, if the safe crop area extends outside the Visible Rendering Region for TTML content in any dimension then the presentation processor shall apply translation and/or scaling to the root container region so that, after this transformation, the Visible Rendering Region does fully enclose the safe crop area.

Note that translation and/or scaling of the root container region effectively changes the origin and extent of the Visible Rendering Region when expressed in the coordinates of the document instance, since the actual rendering area corresponding to the Visible Rendering Region is assumed to be held constant.

Processors are required to maintain the position of the root container region, and by extension each region, relative to the Related Video Object however in some cases display systems can be configured not to show the entire video frame, so the processor needs to make a decision on how to present the subtitle content, with the typical options being to scale or crop.

For example if a 4:3 video is being displayed on a 16:9 display the system could display only the central 16:9 portion of that video, by cropping the top and bottom of the 4:3 video frame. This would be a Visible Rendering Region whose effective origin is 0% 12.5% and whose effective extent is 100% 75%. If the safe crop area were to extend beyond that Visible Rendering Region, e.g. a 14:9 full height horizontally centered area, the content would be cropped. After applying suitable translation and/or scaling to the video and the root container region the resultant Visible Rendering Region fully encloses (inclusively of edges) the safe crop area. The following diagram illustrates this example:



Note: the video presentation system behaviour is out of scope of this document, however it would be possible for the video to be translated and/or scaled alongside the root container region to ensure that the relative positions of TTML content and video are maintained; this behaviour could be a user configuration setting.

Note: there are already conformance requirements for processors in how they relate the root container region to the related video object. If the safe crop area does not extend outside the Visible Rendering Region for TTML content then those conformance requirements suffice to ensure that the presentation processor displays that part of the root container region that overlaps with the Visible Rendering Region, maintaining relative positioning with any related video object.

Note: the translation and scaling algorithm is left undefined and is therefore implementation specific, or may be specified by the interchange or processing context.

Note: scaling algorithms should take into account that text size scales with height not width; anamorphic scaling could result in text occupying an unexpected width, leading to overflow and wrap that were not in the original authored document.

Impact on subtitle document authors

Subtitle document authors should not rely on overflow behaviour, and should create regions that are big enough to contain the text that is flowed into them at all points during the presentation. Since overflow could result in text being rendered outside its region, that text could be cropped unintentionally without any way to signal to presentation processors that it should be rendered fully.

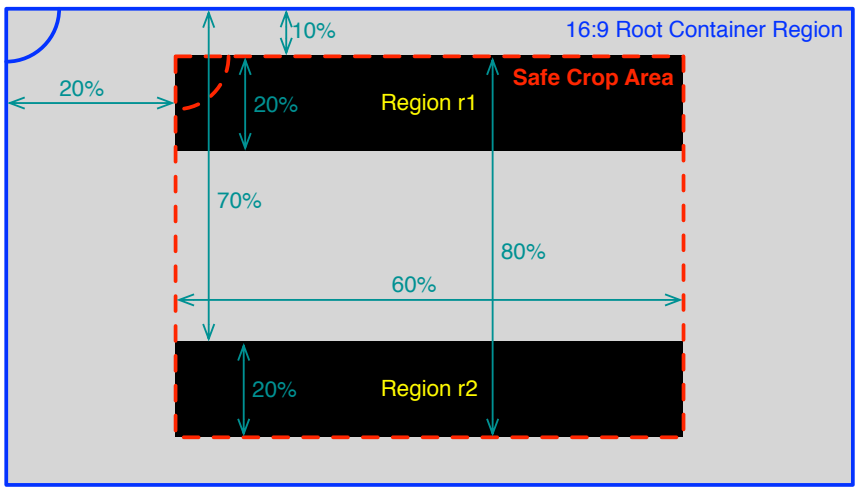
Note: there are no required behaviours of presentation processors in overflow scenarios; clipping text when overflow is indicated by the document instance is not recommended but does not prevent a processor from being conformant.

Examples

Example 1

The following example document fragment authored for 16:9 video illustrates the concept. Note that the `ittp:aspectRatio` parameter is included here to define the aspect ratio of the root container region, which is a related but orthogonal concept to the safe crop area. Some interchange and processing contexts do not require the root container region aspect ratio to be signalled, for example where the aspect ratio is required to be identical to the aspect ratio of the decoded video.

```
<tt
  xmlns="http://www.w3.org/ns/ttml"
  xmlns:ttp="http://www.w3.org/ns/ttml#parameter"
  xmlns:ittp="http://www.w3.org/ns/ttml/profile/imsc1#parameter"
  xmlns:[namespace]="[namespace URN]"
  ittp:aspectRatio="16 9"
  [namespace]:safeCropArea="20% 10% 60% 80%" >
    <head>
        <layout>
            <region xml:id="r1" tts:origin="20% 10%" tts:extent="60% 20%"/>
            <region xml:id="r2" tts:origin="20% 70%" tts:extent="60% 20%"/>
        </layout>
    </head>
    <body>
        <div>
            <p begin="00:00:01" end="00:00:02" region="r1">This line in region r1</p>
            <p begin="00:00:10" end="00:00:11" region="r2">This line in region r2</p>
        </div>
    </body>
</tt>
```

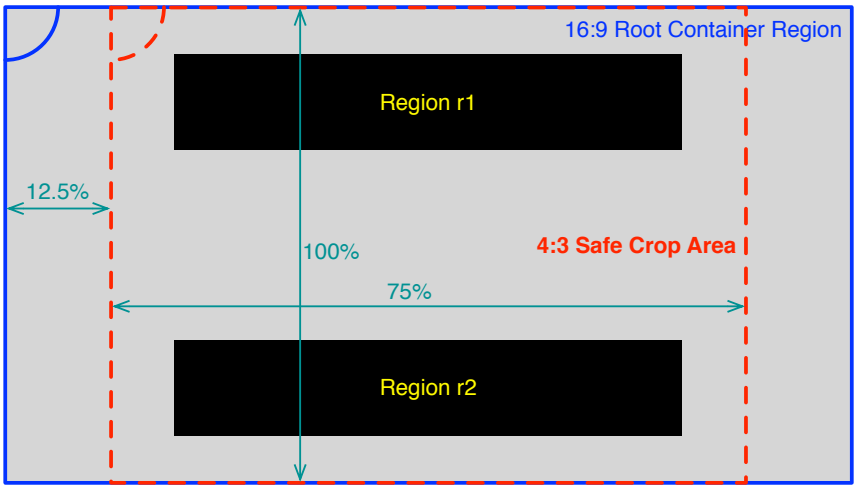


The above example defines two regions that are both referenced. The smallest bounding box enclosing both has a top left 20% in from the left and 10% down from the top, and a size 60% wide and 80% high. This bounding box is used as the value for the `safeCropArea` parameter.

Example 2

A value of `safeCropArea` of `"75% 100%"` within a 16:9 root container region would be the equivalent of a 4:3 area centred horizontally. Such a value could be used to signal to a processor that when displaying 16:9 video on a 4:3 display using centre cut-out no translation or scaling of the subtitle area is needed.

```
<tt
  xmlns="http://www.w3.org/ns/ttml"
  xmlns:ttp="http://www.w3.org/ns/ttml#parameter"
  xmlns:[namespace]="[namespace URN]"
  [namespace]:safeCropArea="12.5% 0% 75% 100%" >
...
</tt>
```

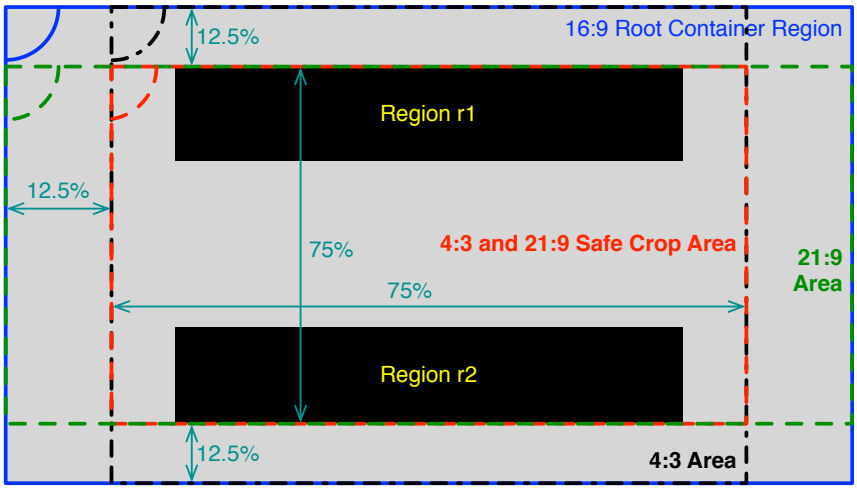


In this example the 12.5% 0% origin is calculated by centring the safe crop area relative to the root container region.

Example 3

A value of `safeCropArea` of `"75% 75%"` within a 16:9 root container region would have the equivalent of a 4:3 area's width centred horizontally and a 21:9 area's height centred vertically[1]. Such a value could be used to signal to a processor that when displaying 16:9 video on either a 4:3 display or a 21:9 using centre cut-out no translation or scaling of the subtitle area is needed.

```
<tt
  xmlns="http://www.w3.org/ns/ttml"
  xmlns:ttp="http://www.w3.org/ns/ttml#parameter"
  xmlns:[namespace]="[namespace URN]"
  [namespace]:safeCropArea="12.5% 12.5% 75% 75%" >
    <head>
        <layout>
            <region xml:id="r1" tts:origin="20% 12.5%" tts:extent="60% 20%"/>
            <region xml:id="r2" tts:origin="20% 68.5%" tts:extent="60% 20%"/>
        </layout>
    </head>
    <body>
        <div>
            <p begin="00:00:01" end="00:00:02" region="r1">This line in region r1</p>
            <p begin="00:00:10" end="00:00:11" region="r2">This line in region r2</p>
        </div>
    </body>
</tt>
```



This example is not necessarily desirable in all cases, but is included to demonstrate that the proposed approach at least permits a safe area to be defined that is not 100% of the height or the width of the root container region. It is common for at least part of the lower or upper 12.5% to be used for some subtitle or caption content, but if a content provider wishes to ensure that subtitles can be provided that are safe in these scenarios then this example would be effective.

An alternative reasonable approach could be to specify a height of, say 84% and to limit in the video presentation system the scaling of the related video object so that on a 21:9 display narrow black bars are visible left and right while displaying all of the Safe Crop Area.

[1] 21:9 is commonly used as a mathematically inaccurate synonym for 64:27. A 21:9 area the same width as a 16:9 area would occupy 76.2% of the height, whereas a 64:27 area would occupy 75% of the height.