International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.302
Amendment 1
(01/2019)

SERIES J: CABLE NETWORKS AND TRANSMISSION OF TELEVISION, SOUND PROGRAMME AND OTHER MULTIMEDIA SIGNALS

Application for Interactive Digital Television – Part 2

System specifications of augmented reality smart television service

Amendment 1

Recommendation ITU-T J.302 (2016) – Amendment 1

International Telecommunication Union

Recommendation ITU-T J.302

System specifications of augmented reality smart television service

Amendment 1

Summary

Augmented reality (AR) technology is a kind of mixed reality in which 2D/3D graphics are integrated into the real world in order to enhance user experience and enrich information. Augmented reality smart television (AR-STV) is TV broadcasting using AR technology to blend augmented content (2D/3D graphic object) with broadcasting content in real-time on receiving terminals. The augmented content overlaid in the augmentation region is fixed statically or moves along the trajectory as if an AR-based location application shows path information on the real location from a smart-phone's camera. AR-STV provides that TV viewers can choose whether to turn on the augmented content or watch the original TV content only (without the augmented content).

Recommendation ITU-T J.302 specifies the related technologies that should be implemented for an augmented reality smart television system.

Amendment 1 introduces support for real-time comment sharing services.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T J.302	2016-10-14	9	11.1002/1000/13054
1.1	ITU-T J.302 (2016) Amd. 1	2019-01-13	9	11.1002/1000/13835

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Recommendation ITU-T J.302

System specifications of augmented reality smart television service

Amendment 1

Editorial note: This is a complete-text publication. Modifications introduced by this amendment are shown in revision marks relative to Recommendation ITU-T J.302 (2016).

1 Scope

Figure 1 shows the reference architecture of the augmented reality smart television (AR-STV) system.

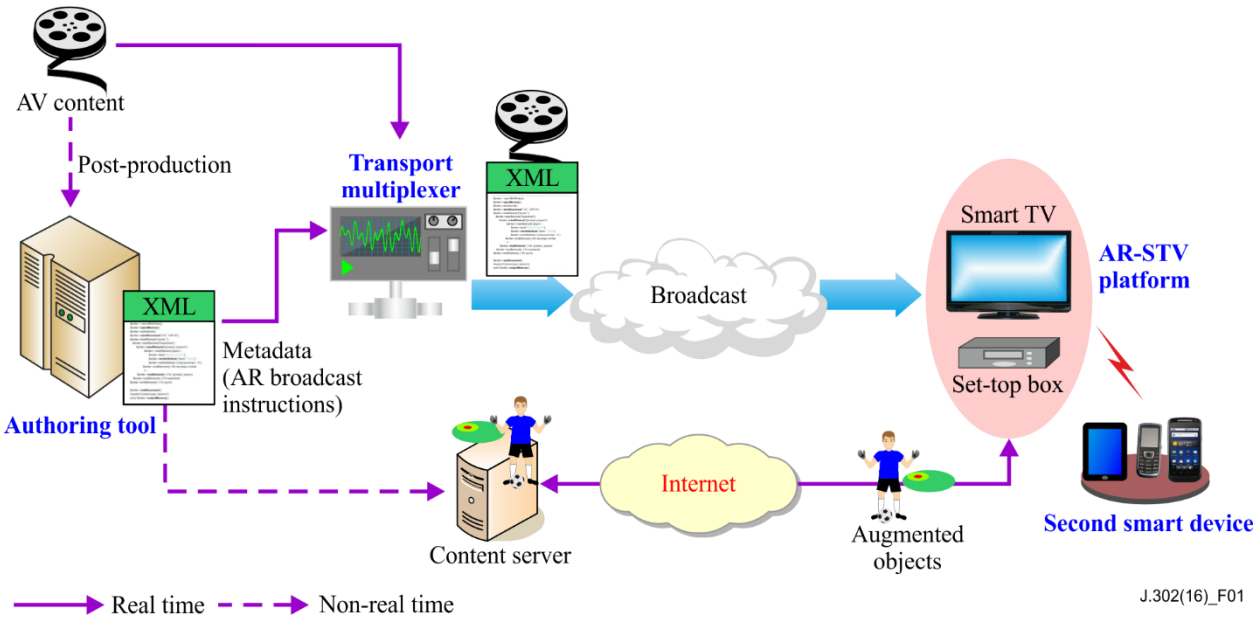


Figure 1 – System architecture of AR-STV

Augmented reality smart television (AR-STV) provides an AR-based TV broadcasting service to implement the mixed content of a broadcasting programme and augmented content at a targeted position (augmentation region) and time (augmentation time) in real-time on a receiving terminal such as a smart TV or set-top-box (STB). AR-STV overlays augmented contents provided by the broadcasting provider itself or by the content providers which are authorised by the broadcasting provider. When AR-STV is provided, the broadcasting signal includes the metadata which contains the permissions about timing and the area for the overlay. Authorisation information for an AR content provider is also provided in the broadcasting signal. This mechanism does not permit any illegal overlay.

The AR-STV is realized by the combination of a specific signal on the broadcasting programme, metadata, augmented content and terminal side synchronization. The terminal functionality can be implemented as the extension of the existing integrated broadcast-broadband (IBB) systems. The scope of this Recommendation comprises the following technologies to realize AR-STV:

– Synchronization scheme for AR-STV
– Metadata for AR-STV
– Signalling protocol for AR-STV

This Recommendation specifies three technologies for AR-STV, (1) synchronized representation between a broadcasting programme and augmented content, (2) metadata and (3) signalling protocol according to functional requirements defined in [ITU-T J.301].

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T J.301] Recommendation ITU-T J.301 (2014), *Requirements for augmented reality smart television systems*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 augmented broadcasting [ITU-T J.301]: Broadcasting service or programme to realize augmented reality smart television (AR-STV).

3.1.2 augmented content [ITU-T J.301]: A binary object, such as 2D images, 3D animated models or audio/video streaming files, to be augmented into a predefined augmentation region.

3.1.3 augmented reality (AR) [ITU-T J.301]: A type of mixed reality where graphical elements are integrated into the real world in order to enhance user experience and enrich information.

3.1.4 augmented reality smart television (AR-STV) [ITU-T J.301]: AR-based TV broadcasting service to implement the mixed content of a broadcast programme and augmented object at a targeted position and time in real time on a receiving terminal such as a TV or set-top box.

3.1.5 augmentation region [ITU-T J.301]: The targeted area to augment any object or region in a TV scene. It provides the parameters of a static or animated 2D region related to natural media.

3.1.6 augmentation time [ITU-T J.301]: The appointed time to augment any object or region in a TV scene. It provides the parameters of start presentation time and time duration for augmented content.

3.2 Terms defined in this Recommendation

This Recommendation defines the following term:

3.2.1 reference image: A resource to be referenced in the tracking process of acquiring coordinates generated by a particular region's movement in the video scene.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ABM Augmented Broadcasting Metadata

APP Application

AR Augmented Reality

AR-STV Augmented Reality Smart Television

EIT Event Information Table

FoV Field of View

IBB Integrated Broadcast-Broadband

MPEG Moving Picture Experts Group

MUX Multiplexer

PID Programme Identification

PMT Program Map Table

PSI Program Specific Information

PSIP Program and System Information Protocol

STB Set-Top-Box

TS Transport Stream

VCT Virtual Channel Table

5 Conventions

In this Recommendation:

The keywords "is required to" indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this document is to be claimed.

The keywords "is recommended" indicate a requirement which is recommended but which is not absolutely required. Thus this requirement need not be present to claim conformance.

The keywords "is prohibited from" indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this document is to be claimed.

The keywords "can optionally" indicate an optional requirement which is permissible, without implying any sense of being recommended. This term is not intended to imply that the vendor's implementation must provide the option and the feature can be optionally enabled by the network operator/service provider. Rather, it means the vendor may optionally provide the feature and still claim conformance with the specification.

In the body of this document and its annexes, the words *shall*, *shall not*, *should* and *may* sometimes appear, in which case they are to be interpreted, respectively, as *is required to*, *is prohibited from*, *is recommended* and *can optionally*. The appearance of such phrases or keywords in an appendix or in material explicitly marked as *informative* are to be interpreted as having no normative intent.

6 Signalling protocol for AR-STV

The real-time broadcast stream to be provided to the receiving terminal should include signalling and metadata for the AR-STV service. The receiving terminal may determine whether to provide an AR-STV service by checking an AR-STV descriptor contained in the program initialization information of the broadcast stream. Table 1 illustrates an example of the syntax of an AR-STV descriptor including descriptor_tag, descriptor_length and augmented_broadcasting_service_type. More specifically, the descriptor_tag is information that identifies an associated descriptor as an AR-STV descriptor. Descriptor_length indicates the total length of the descriptor. The ARSTV_app_type is to distinguish application types of AR-STV service. ContentProvider_Info includes content providers' information. The descriptor may be located in a table of a section packet, such as a virtual channel table (VCT) or event information table (EIT) of program and system information protocol (PSIP) and/or a program map table (PMT) of program specific information (PSI) in MPEG-2 TS. Table 1 shows the ARSTV descriptor syntax.

Table 1 – ARSTV descriptor syntax

Syntax	Semantics
ARSTV_descriptor() {	
Descriptor_tag	Descriptor identifier
Descriptor_length	Descriptor length
ARSTV_app_type	Application Service Type
ContentProvider_Info {	

ContentProvider (N)
For(i=0;i<N;i++) {
CP_Code
CP_Name
CP_Service (M)
For(j=0;j<M;j++) {
Service_Name
CP_URL
}
}
	Content Provider Count
	
	Content Provider indication
	Print name
	Service count
	
	Service name
	Service URL

}	
}	

In Table 2, which shows the metadata PES packet payload, the payload of metadata packet includes an identifier to indicate metadata sequence. An Initial_program_reference_clock includes synchronization information for indicating a start point of a currently broadcast programme. The 33-bit information of the Initial_program_reference_clock is Program_Clock_Reference (PCR) information of an MPEG-2 system, corresponding to PCR_base in the existing PCR time and is reference clock information required to calculate synchronization time. ABM_markup_type defines a file format of the metadata. For example, "01" may represent XML format and "10" may represent binary format. ABM_delivery_type_flag is a field that determines whether to receive the metadata over the broadcasting network or over broadband.

Table 2 – Metadata PES packet payload

Syntax	Semantics
ABM_PES_packet_data_byte() {	
Identifier	data identifier
Initial_Program_reference_clock	starting point in broadcasting stream
ABM_markup_type	Metadata file format
ABM_delivery_type_flag	Metadata delivery network type

if(ABM_delivery_type_flag == 0)
{
ABM_data_length
ABM_data_byte
}
	Metadata length and bytes in broadcasting network

Table 2 – Metadata PES packet payload

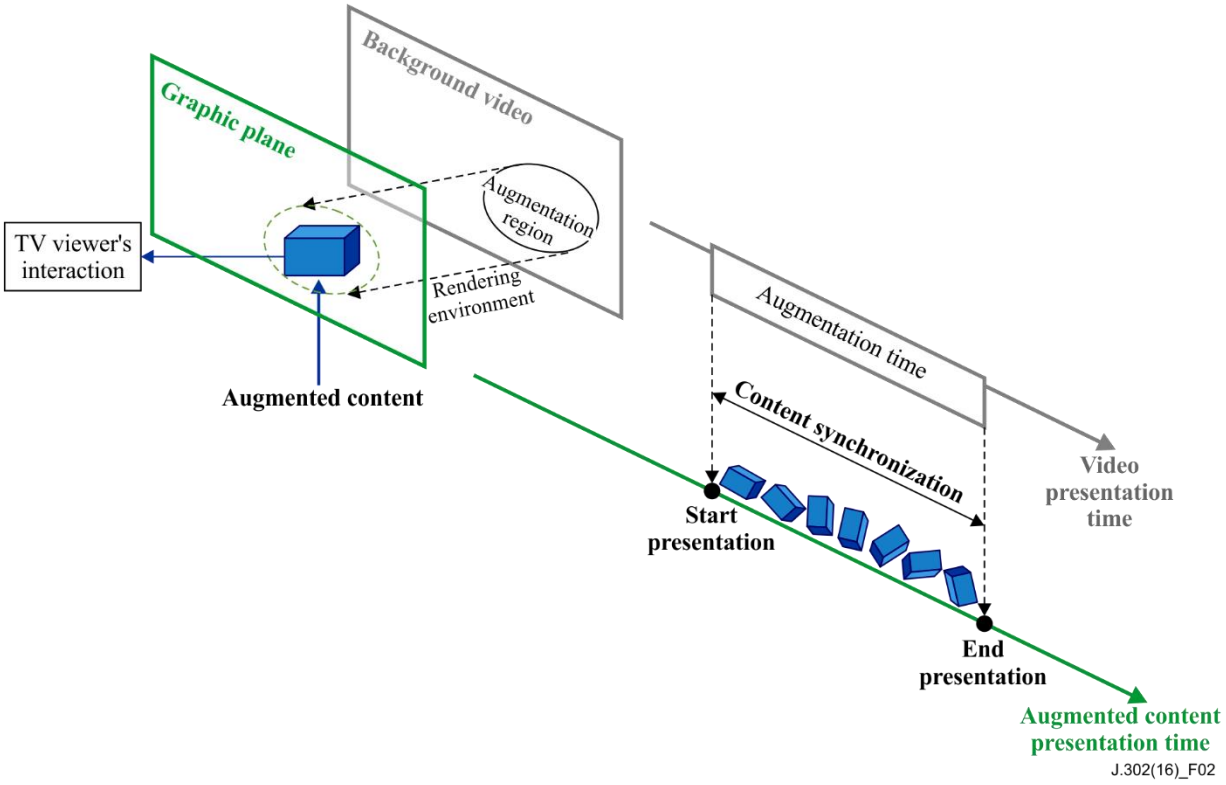
Syntax	Semantics

if(ABM_delivery_type_flag == 1)
{
ABM_URL_length
ABM_URL_byte
}
	Metadata file download information and bytes in broadband network

}	

7 Synchronization scheme for AR-STV

Figure 2 shows the augmentation region, augmentation time, its related attributes and their relationships.



(source: [ITU-T J.301])

Figure 2 – Augmentation region and augmentation time

The AR-STV system presents augmented contents within augmentation regions at the right time on a TV programme. The augmented contents are shown by performing instructions predefined in metadata. Also as shown in Figure 2, each augmented content appears at the appointed time and disappears after finishing rendering during the augmentation time on a TV screen. If augmented contents allow user interaction, the TV viewer can use the interactive service (e.g., the interactive application defined in Appendix I.2 of [ITU-T J.301]) if he/she wants.

The AR-STV system utilizes an initial program clock reference (PCR) as synchronization information with reference to the start point of the broadcast and transmits with metadata through a TS re-multiplexer which re-multiplexes the MPEG-TS stream into a form suitable for AR-STV service. The metadata includes appointed presentation time information (hour:minute:second:frame) which

consists of the elapsed time from broadcasting start and the number of frames. An AR-STV receiving terminal estimates the augmentation time by calculating values with the initial PCR and appointed presentation time referenced in metadata transport. The transmission system for AR-STV is shown in Figure 3.

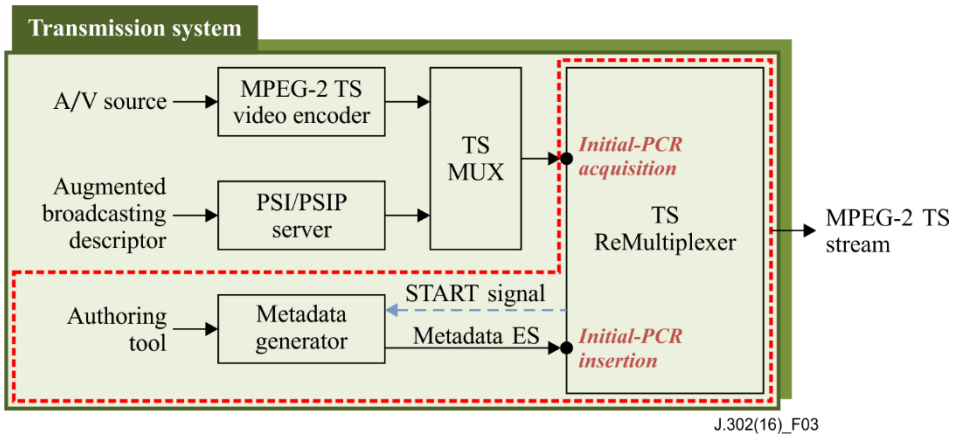


Figure 3 – Structure of transmission system for AR-STV

In Figure 3 the transmission system for AR-STV transmits the AR-STV descriptor as well as metadata that contain the program initialization information described with reference to Table 2 in clause 6. The transport stream (TS) generated from a TS multiplexer (MUX) is transmitted to a TS re-multiplexer and the re-multiplexer re-multiplexes the TS into a form suitable for the AR-STV service. A metadata generator generates metadata expressed as XML data or a TS stream encoding XML data and inputs the generated metadata to the re-multiplexer. The re-multiplexer multiplexes the metadata with the TS and transmits the resulting data to a receiving terminal. The re-multiplexer multiplexes the metadata according to a syntax described with reference to Table 2. In one example, the metadata generator stores the initial PCR, i.e., the Initial_Program_reference_clock of Table 2, in the re-multiplexer as synchronization information with reference to the start point of the broadcast programme as soon as a start signal from the re-multiplexer is identified; in addition, the re-multiplexer multiplexes the metadata, which are arranged based on the transmission time to be transmitted at a particular time, with the stored initial PCR, in units of PES and transmits the metadata at the time when a null packet is detected during multiplexing TS streams. The PSIP generator creates an augmented broadcasting descriptor that describes the augmented broadcasting and inputs it to the re-multiplexer.

When an MPEG-2 TS stream including a broadcast programme for an AR-STV service is emitted in real time, the re-multiplexer continuously reads each individual TS packet unit of the MPEG-2 TS stream. At this time, the value of the initial PCR of the input MPEG-2 packet as timing information for content synchronization is stored in a memory. The re-multiplexer multiplexes the metadata with the stored initial PCR in units of PES of metadata and detects whether there are metadata to be transmitted in real time during this process. If there are metadata to be transmitted at a particular time, the re-multiplexer efficiently executes transmission without additional bandwidth load by carrying the metadata instead of the null packet. As the metadata has properties of randomly ordered streams, rather than consecutive data streams of the broadcast programme, the timing information for the augmented broadcasting synchronization is transmitted only when the metadata is sent and thus the receiving terminal does not need to refer to other broadcasting or data-packet information when receiving and parsing the metadata.

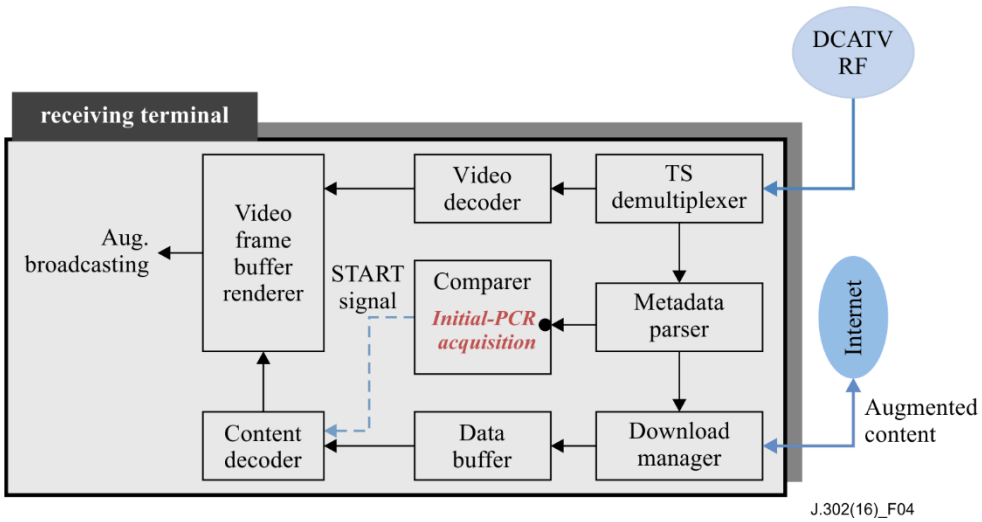


Figure 4 – Structure of receiving terminal for AR-STV

A receiving terminal, as shown in Figure 4, interprets the ARSTV_descriptor contained in the TS to determine whether to receive an augmented broadcast or not. In addition, when the receiving terminal receives the augmented broadcasting metadata packet before the related augmented content and it is completely downloaded, it extracts the initial PCR information for synchronization and XML data for augmented broadcasting from the metadata PES. The extracted XML data are buffered in the data buffer. As an example, a timing comparer calculates the presentation time stamp, which is an MPEG-2 system time clock, in consideration of the current status of the program, using the initial PCR and the activating augmentation time (hour:minute:second:frame) contained in the XML data. The PTS may be calculated using the equation below.

PTS for activating the augmentation time = Initial PCR base + {number of frames converted from the activation time defined by XML data × Interval of PTS (3,000)}

Assuming that the initial PCR is 30,000 and that the programme play time for which content A defined by XML is activated is 1 minute: 10 seconds: 10 frames, the augmented content is activated after 2110 frames. By applying this to the above equation, 2,110 × 3,000 = 6,330,000, as an interval value, is obtained. Then, the PTS value is produced by adding the value of the initial PCR, i.e., 30,000 and the obtained interval value (6,330,000 + 30,000 = 6,360,000).

8 Metadata specification for AR-STV

The metadata for AR-STV delivered with broadcasting content helps the system to overlay graphics on the video. It could be delivered by the broadcasting channel or by the Internet channel separated from the broadcasting stream. The time information that may be associated with the augmentation region (AR) event may be used for the synchronization with respect to the broadcasting contents. The region of the video scene where the graphics shall be overlaid is described with a normalized value according to the resolution of the video. The scale, rotation and transformation of the graphic objects may be updated according to the change of view of the virtual camera. The graphic objects may be embedded in the metadata or indicated with link information. For the synchronized presentation of synthesized graphics with broadcasting content in the AR-STV service, presentation time is very important. Therefore the fragmentation of the metadata will be made by the time point when the graphic object is presented. This unit of fragmentation is called an "instruction" and each instruction contains updated information needs to be presented at a specific time point. Figure 5 shows the streaming strategy of the metadata. The first instruction contains every item of information necessary for the generation of a new AR event such as augmentation region, overlay graphic object and

environmental information. The AR in the first instruction indicates the start position of AR object and this position and orientation may be changed by the following instructions. The following instructions contain only the variation factors such as the change of the position of AR region, the change of the graphic objects, or the change of the illumination direction. In addition the initial instruction enables a terminal to allocate the needed memory and pre-loads augmented objects to the memory space in advance and will be delivered repeatedly in the fixed timing interval within AR-STV service. These strategies will reduce the data overload of the media channel.

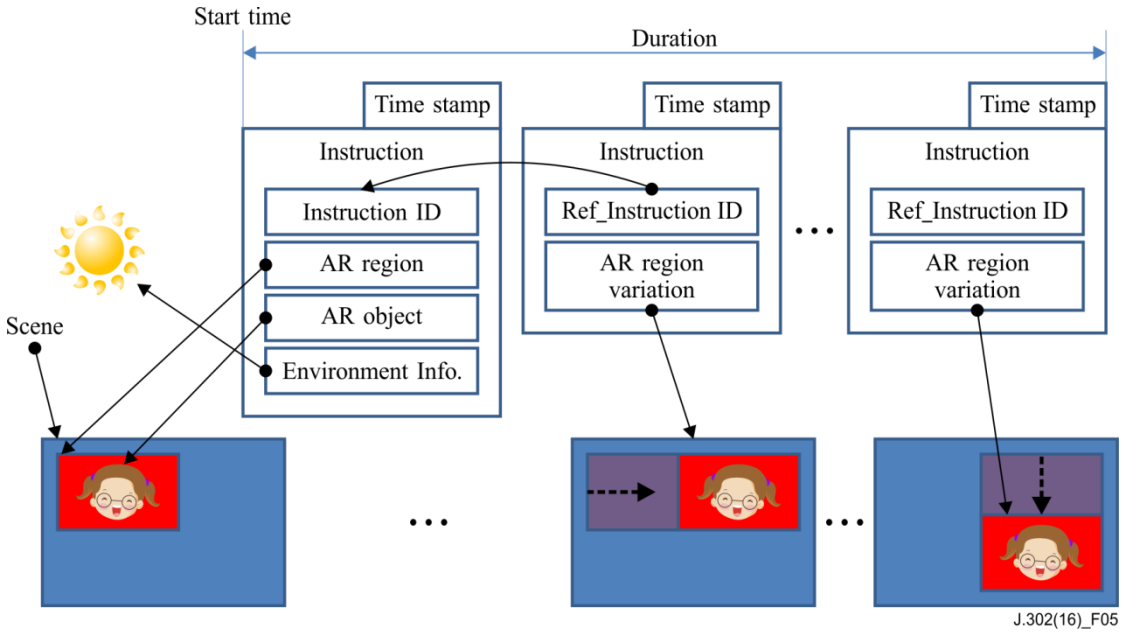


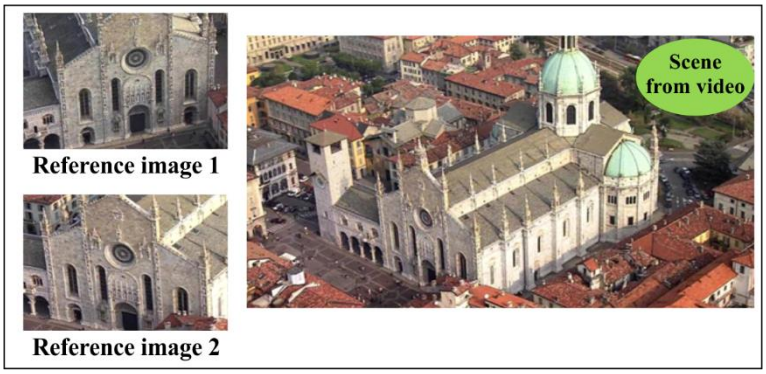
Figure 5 – Streaming strategy for AR-STV metadata

8.1 Initial instruction

Name	Definition	(M)andatory(O)ptional
InitInstruction	Serves as a container for information which shall be transmitted beforehand or periodically.	O
AugmentedObject	Describes the location of an augmented object in order for the terminal to download beforehand.	M
Id	Specifies the ID of an initial instruction.	M
contentsNum	Specifies the number of augmented objects in the initial instruction.	M

8.2 Instruction

Name	Definition	(M)andatory (O)ptional
ReferenceResource	Contains the URL of the image(s) to be tracked from the video scenes in the user terminal. In the case where the user terminal has enough power to process image tracking algorithm, the terminal may overlay 3D graphics on the video updating the motion according to the automatically extracted motion factors from the video. Here 'track' means the moving changes of the image and 'tracking' means a process to acquire coordinates generated by the image's track. Figure 6 shows the reference images for tracking and	O

Name	Definition	(M)andatory (O)ptional
	video scene. Reference images 1 and 2 are examples of resources to be referenced for tracking in the video scene of the church building.  J.302(16)_F06 **Figure 6 – Reference images and scene for tracking from video** To realize the matching between reference image and moving image, several types of tracking technologies can be used. Figure 7 shows the image tracking process by feature point matching algorithm using reference image 1 and 2 in Figure 6.  J.302(16)_F07 **Figure 7 – Process of image tracking algorithm**	
AugmentationRegion	Describes the spatial region to be overlaid with graphics and also the motion of this region so the graphic objects may move according to this motion value. The motion value includes the position, scale and rotation of the object in 3D spaces. In the case where the user terminal does not have enough power for tracking, the metadata should have motion information and the user terminal only performs composition of the graphics. EXAMPLE 1 Scale, rotation, transformation values by x, y and z axis.	M
AugmentedObject	Describes the graphics or texts which will be overlaid on the scene. The graphic or text content may be either embedded in the metadata as inline content, or be referenced by an URL of a remote server.	M
EnvironmentInfo	Describes the environmental elements which will affect natural composition of synthetic graphics on the real world image. EXAMPLE 1 Light source type, direction, colour, etc.	O
UserInteraction	Describes allowed user interaction for the augmentation region.	O

Name	Definition	(M)andatory (O)ptional
firstInstFlag	Indicates that the current instruction is the first instruction containing all information necessary to generate a new AR event. The value 1 means the current instruction is the first instruction and the value 0 means not.	M
augRegionIndicator	Describes the identification number to link the first instruction and the following instructions. By checking this value, terminals can recognize that the instructions belong to the same augmentation region.	M
Pts	Describes the presentation time stamp according to the time scale. The instruction should be ready to be presented up to this time.	M
duration	Describes the time period when the information in the key instruction shall be valid according to the time scale.	M
timeScale	Describes clock tic number per second.	M
numInstruction	Describes the number of the instructions including the key instruction and its following instructions.	M
priority	Describes the priority of instructions in processing order among the instructions sharing the same time point when they should become available for consumption. A value of one indicates the highest priority and larger values indicate lower priorities.	O

8.2.1 Reference resource

Name	Definition	(M)andatory (O)ptional
Resources	Describes the location of reference resource(s) for recognition or tracking algorithm.	M

8.2.2 Augmentation region

Name	Definition	(M)andatory (O)ptional
TransformMaxtrix	Describes the transform matrix for calculating x, y and z point.	O
Coordinate	Describes 3D coordinates for 4 points of the quadrangle representing the augmentation region.	O
X1, y1, z1	Describes left-top coordinate.	M
X2, y2, z2	Describes right-top coordinate.	M
X3, y3, z3	Describes right-bottom coordinate.	M
X4, y4, z4	Describes left-top coordinate.	M
SRT	Describes rotation, scale and translation factor for central point of augmentation region.	O
sx, sy, sz	Scale factor of x, y and z axis.	M
rx, ry, rz	Rotation factor of x, y and z axis.	M
tx, ty, tz	Translation factor of x, y and z axis.	M
* Author should select one of TransformMatrix, Coordinate, or SRT.		

8.3 Augmented object

Name	Definition	(M)andatory (O)ptional
Inline	This parameter describes binary format whether augmented content is included in the metadata or not. Basically, augmented content will be sent separately from the metadata. However, if the system sends the augmented content including in the metadata, it is required to set this parameter and indicates augmented content embedded in metadata.	O
Remote	Describes the URI indicating remote augmented content.	O
Tactile	This parameter describes the binary format of tactile content embedded in metadata and will be used when haptic device is connected to the AR-STV terminal. Appendix II shows more detail about a haptic device use case.	O
ArrayIntensity	Describes the intensity of actuator by n*n matrix. It indicates how much space is applied for the tactile material.	M
tactileEffect	Describes the type of tactile actuator. For example; pressure, vibration	M
timeSamples	Describes the number of samples which shall be updated per second.	M
clearFlag	Indicating whether the previous augmented content shall be clear or not before displaying new augmented content. Value 1 means the previous content should be cleared.	O
Service	Describes the service type of augmented broadcasting. This parameter is for future use.	O
* Author should select one of Inline, Remote and Tactile		

8.4 Environment info

8.4.1 Semantics

Name	Definition	(M)andatory (O)ptional
GlobalPosition	Describes GPS information of augmentation region.	O
Address	Describes the physical address indicated by augmentation region.	O
longitude	Describes the longitude of augmentation region.	M
Latitude	Describes the latitude of augmentation region.	M
Light	Describes the light source for natural composition of augmented content.	O
Position	Describes the location of light source with 3D position (x, y, z).	O
Rotation	Describes the location of light source with 3D rotation (x, y, z).	O
Type	Describes the type of light source according to below value and meaning; 1: point light 2: directional light 3: spot light	O

Name	Definition	(M)andatory (O)ptional
Colour	Describes the colour of light source with RGB format. Ex) #FF0000	O
intensity	Describes the intensity of light source.	O
Camera	Describes the camera characteristics. This parameter will be used by the rendering engine to generate a synthesized video frame.	O
Fov	Describes the field of view (FoV).	O

8.5 User interaction

Name	Definition	(M)andatory (O)ptional
`Resource`	Describes URI of graphic object which will replace the current augmented object.	O
`SRT`	Describes the updated position and orientation and scale of the augmented object.	O
`Event`	Describes the event from user. The replacement, position change, scale change of the augmented object shall happen according to the event value.	O

Annex A

Real-time comment sharing

(This annex forms an integral part of this Recommendation.)

Many online video services, especially those for comic videos, provide the "bullet screen" or "danmu" feature which consists of real-time text comments flying consecutively across the screen (e.g., from right to left) from people watching the video at the same time. When a person is watching a video, he or she can also view comments from the others as well as share his or her own comments with other viewers. This feature is a tool for socially viewing videos. It is distinctive as it is real-time. The basic features of danmu are as follows:

1) Basically, the danmu flies horizontally from the right side to the left side of the screen.

2) After the danmu enters the screen, it moves according to a certain trajectory.

3) Danmu movement speed is determined by the content length (the longer the content, the faster it moves).

A.1 Danmu reference model

Figure A.1 shows a reference model for the danmu function that complies with the reference architecture of Figure 1.

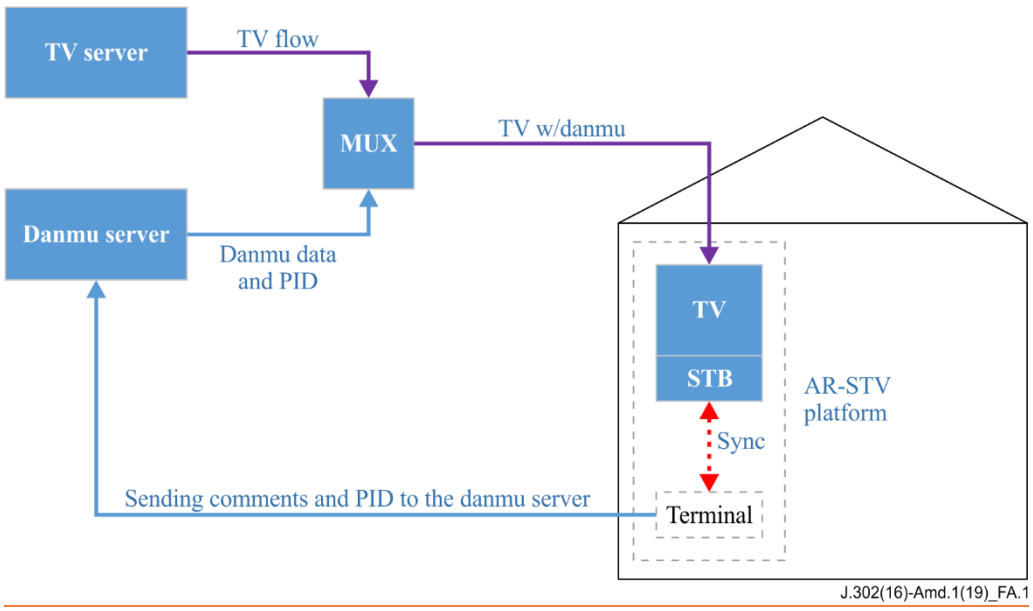


Figure A.1 – Reference model for the danmu function

In Figure A.1, the terminal is for commenting. It can have various forms. Here are three examples:

- A soft keyboard showing on the TV screen with a remote controller;
- A soft keyboard embedded in an APP on a cell phone connected or synchronized to the STB or TV server;
- A hard keyboard with synchronization to the STB or TV server.

Besides typing comments, the terminal has the capability to transmit those comments to the danmu server. The transmission of the comments can either use the existing cable TV network or an Internet connection. The comments shall be transmitted to the dammu server together with the programme information (e.g., PID).

The AR-STV platform is comprised of the TV, STB and the (embedded) terminal. Other than receiving data, this AR-STV platform has the capability to transmit comments (e.g., text, emoji, etc.) to a specific server.

A.3 Danmu metadata specification

The danmu is an augmented object. It shall use the same specification as given in clause 8 for instructions and augmented object descriptions.

A.4 Danmu synchronization

The danmu function shall use the scheme specified in clause 7 for synchronization.

A.5 Danmu mechanism

The danmu mechanism shall use the following procedure:

Step 1. After a customer types comments and the danmu APP sends them to a danmu server, the comments shall be either stored in a comments queue in the danmu server, or be sent to a MUX device directly if the queue is empty. The MUX device is responsible for the multiplexing of the video traffic and danmu comments. As shown in the example of Figure A.2, six danmu tracks are stored in the danmu system. The danmu server will conduct the track that each content goes into.

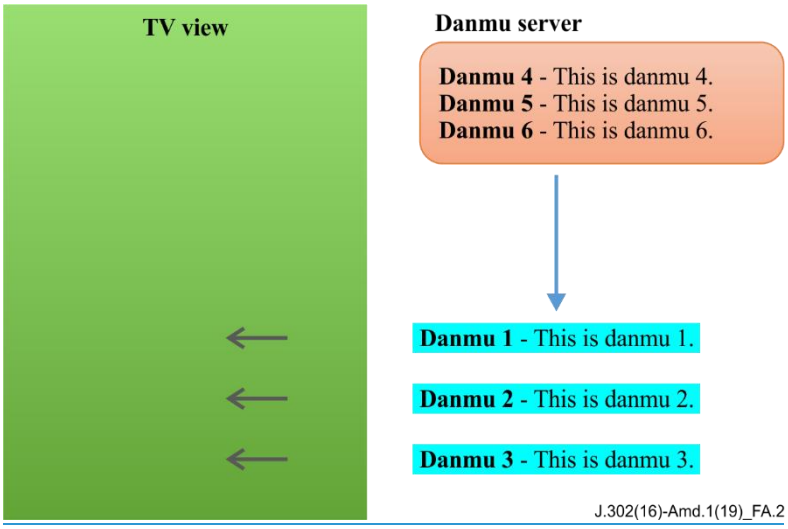


Figure A.2 – Danmu process

Step 2. The danmu comments shall fly horizontally across the screen from right to left at a speed that should be determined according to the length of the content.

NOTE – Determination of movement speed is vendor discretionary.

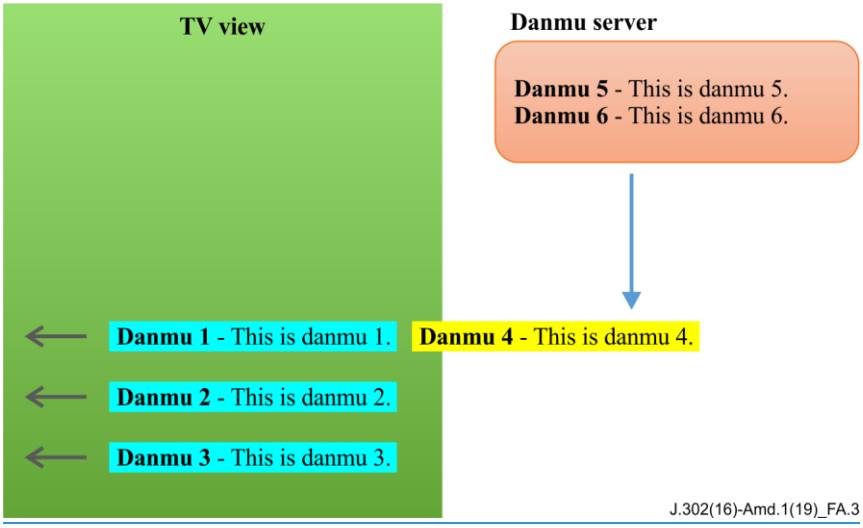


Figure A.3 – Danmu contents display

Step 3. The track should be deleted from the danmu system once the danmu content has completely finished flying across the screen. See Figure A.4:

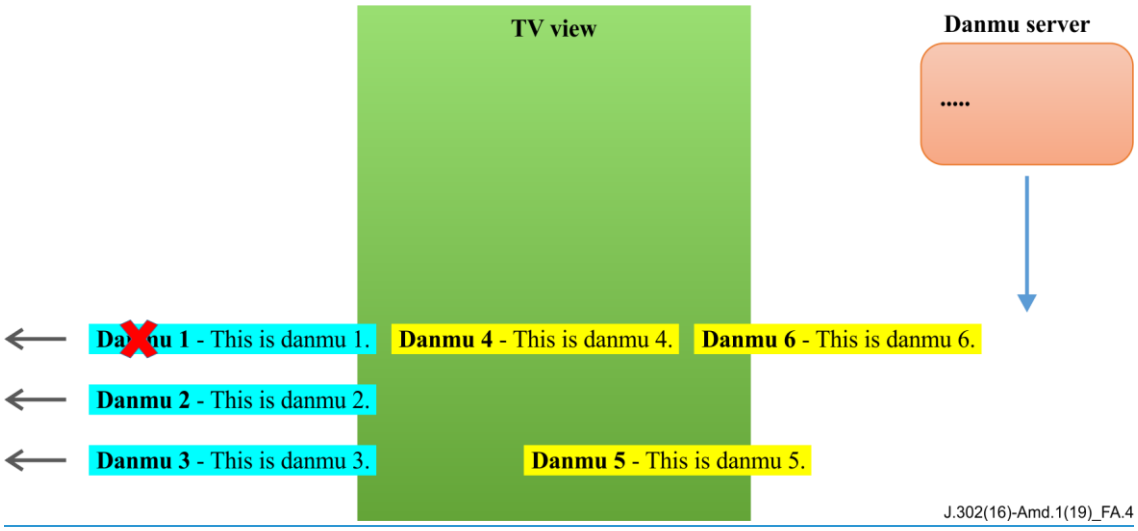


Figure A.4 – Danmu contents display

Appendix I

Syntax of XML-based AR-STV metadata

(This appendix does not form an integral part of this Recommendation.)

This appendix shows syntax of XML-based AR-STV metadata according to clause 8.

I.1 Namespace

The prefix and the namespace of the AR-STV metadata is shown in Table I.1.

Table I.1 – Prefixes and namespace

Prefix	Corresponding namespace
ARSTV	urn:arstv:ver1:represent:ARSTVmetadata:2011:07

I.2 Schema wrapper

In order to form a valid schema document, these schema components should be gathered in a single document with the following declaration defining in particular the target namespace and the namespaces prefixes.

```
<?xml version="1.0"?>
<?xml version="1.0" encoding="UTF-8"?>
<schema xmlns:ARSTV="urn:arstv:ver1:represent:ARSTVmetadata:2011:07"
xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
xmlns="http://www.w3.org/2001/XMLSchema"
targetNamespace="urn:arstv:ver1:represent:ARSTVmetadata:2011:07"
elementFormDefault="qualified" attributeFormDefault="unqualified">
      <import                      namespace="urn:mpeg:mpeg7:schema:2001"
schemaLocation="mpeg7-cdp-2004.xsd"/>
```

Additionally, the following line should be appended to the resulting schema document in order to obtain a well-formed XML document.

```
</schema>
```

I.3 Root element

```
<!-- ################################################ -->
<element name="ARSTV" type="ARSTV:ARSTVType"/>
<complexType name="ARSTVType">
  <sequence>
    <element name="DescriptionMetadata"
type="mpeg7:DescriptionMetadataType" minOccurs="0"/>
    <element name="InitialInstruction" type="abm:InitialInstructionType"
minOccurs="0" maxOccurs="unbounded"/>
    <element name="Instruction" type="ARSTM:InstructionType"
minOccurs="0" maxOccurs="unbounded"/>
  </sequence>
</complexType>
```

I.4 Initial instruction

```
<!-- ################################################ -->
<!-- Inital Instruction type                  -->
<!-- ################################################ -->
```

```
<element name="InitInstruction" type="arstv:InitialInstructionType"/>
  <complexType name="InitialInstructionType">
    <sequence>
      <element name="AugmentedObject" type="arstv:AugmentedObjectType"
maxOccurs="unbounded"/>
    </sequence>
    <attribute name="id" type="ID" use="optional"/>
    <attribute name="contentsNum" type="unsignedInt" use="optional"/>
</complexType>
```

I.5 Instruction

```
<!-- ################################################ -->
<!-- Instruction Base type                            -->
<!-- ################################################ -->
<complexType name="InstructionBaseType" abstract="true">
  <complexContent>
    <restriction base="anyType">
      <attribute name="id" type="ID" use="optional"/>
    </restriction>
  </complexContent>
</complexType>
<!-- ################################################ -->
<!-- Instruction type                                 -->
<!-- ################################################ -->
<complexType name="InstructionType">
  <complexContent>
    <extension base="ARSTV:InstructionBaseType">
      <sequence>
        <element name="ReferenceResources"
type="arstv:ReferenceResourcesType" minOccurs="0"/>
        <element name="AugmentationRegion" type="
arstv:AugmentationRegionType" />
        <element name="AugmentedObject" type=" arstv:AugmentedObjectType"
/>
        <element name="EnvironmentInfo" type=" arstv:EnvironmentInfoType"
minOccurs="0" maxOccurs="unbounded"/>
        <element name="UserInteraction" type=" arstv:UserInteractionType"
minOccurs="0" maxOccurs="unbounded"/>
      </sequence>
      <attribute name="firstInstFlag" type="boolean" use="required"/>
      <attribute name="augRegionIndicator" type="unsignedInt"
use="required"/>
      <attribute name="pts" type="unsignedInt" use="required"/>
      <attribute name="duration" type="unsignedInt" use="required"/>
      <attribute name="timeScale" type="unsignedInt" use="required"/>
      <attribute name="numInstruction" type="unsignedInt" use="required"/>
      <attribute name="priority" type="unsignedInt" use="required"/>
    </extension>
  </complexContent>
</complexType>
```

I.6 Reference resource

```
<!-- ######################################### -->
<!-- Definition of Reference Resource Type   -->
<!-- ######################################### -->
<complexType name="ReferenceResourcesType">
```

```
    <sequence>
       <element name="Resources" type="string" maxOccurs="unbounded"/>
    </sequence>
</complexType>
```

I.7 Augmentation region

```
<!-- ########################################## -->
<!-- Definition of Augmentation Region Type   -->
<!-- ########################################## -->
<complexType name="AugmentationRegionType">
  <sequence>
    <element     name="TransformMatrix"     type="ARSTV:FloatMatrixType"
minOccurs="0"/>
    <element        name="Coordinates"        type="ARSTV:CoordinateType"
minOccurs="0"/>
    <element name="SRT" type="ARSTV:SRTType" minOccurs="0"/>
  </sequence>
</complexType>
<complexType name="CoordinateType">
    <attribute name="x1" type="ARSTV:zeroToOneType" use="required"/>
    <attribute name="y1" type="ARSTV:zeroToOneType" use="required"/>
    <attribute name="z1" type="ARSTV:minusOneToOneType" use="required"/>
    <attribute name="x2" type="ARSTV:zeroToOneType" use="required"/>
    <attribute name="y2" type="ARSTV:zeroToOneType" use="required"/>
    <attribute name="z2" type="ARSTV:minusOneToOneType" use="required"/>
    <attribute name="x3" type="ARSTV:zeroToOneType" use="required"/>
    <attribute name="y3" type="ARSTV:zeroToOneType" use="required"/>
    <attribute name="z3" type="ARSTV:minusOneToOneType" use="required"/>
    <attribute name="x4" type="ARSTV:zeroToOneType" use="required"/>
    <attribute name="y4" type="ARSTV:zeroToOneType" use="required"/>
    <attribute name="z4" type="ARSTV:minusOneToOneType" use="required"/>
</complexType>
<complexType name="SRTType">
    <attribute name="sx" type="float" use="required"/>
    <attribute name="sy" type="float" use="required"/>
    <attribute name="sz" type="float" use="required"/>
    <attribute name="rx" type="float" use="required"/>
    <attribute name="ry" type="float" use="required"/>
    <attribute name="rz" type="float" use="required"/>
    <attribute name="tx" type="float" use="required"/>
    <attribute name="ty" type="float" use="required"/>
    <attribute name="tz" type="float" use="required"/>
</complexType>
<!-- FloatMatrixType -->
<complexType name="FloatMatrixType">
  <simpleContent>
    <extension base="ARSTV:FloatVector">
      <attribute ref="mpeg7:dim" use="required"/>
    </extension>
  </simpleContent>
</complexType>
<simpleType name="FloatVector">
  <list itemType="float"/>
</simpleType>
<simpleType name="zeroToOneType">
  <restriction base="float">
    <minInclusive value="0.0"/>
```

```
    <maxInclusive value="+1.0"/>
  </restriction>
</simpleType>
<simpleType name="minusOneToOneType">
  <restriction base="float">
    <minInclusive value="-1.0"/>
    <maxInclusive value="+1.0"/>
  </restriction>
</simpleType>
```

I.8 Augmented object

```
<!-- ########################################## -->
<!-- Definition of Augmented Object Type    -->
<!-- ########################################## -->
<complexType name="AugmentedObjectType">
  <choice>
    <element name="Inline" type="mpeg7:InlineMediaType" minOccurs="0"/>
    <element name="Remote" type="anyURI" minOccurs="0"/>
    <element name="Tactile" type="arstv:TactileType" minOccurs="0"/>
  </choice>
  <attribute name="targetID" type="anyURI" use="optional"/>
  <attribute name="clearFlag" type="boolean" use="optional"/>
  <attribute     name="service"     type="unsignedInt"     use="optional"
default="1"/>
    <simpleType>
      <restriction base="string">
        <enumeration value="happy"/>
        <enumeration value="sad"/>
        <enumeration value="angry"/>
        <enumeration value="sick"/>
      </restriction>
    </simpleType>
  </attribute>
</complexType>
<!-- ########################################## -->
<!-- Definition of Tactile Type          -->
<!-- ########################################## -->
<complexType name="TactileType">
  <sequence>
    <element name="ArrayIntensity" type="mpeg7:FloatMatrixType"/>
  </sequence>
  <attribute name="tactileEffect" type="arstv:tactileEffectType"
use="required"/>
  <attribute name="timeSamples" type="positiveInteger" use="required"/>
</complexType>
<simpleType name="tactileEffectType">
    <restriction base="string">
       <enumeration value="pressure"/>
      <enumeration value="vibration"/>
      <enumeration value="electric"/>
    </restriction>
</simpleType
```

I.9 Environment info

```
<!-- ########################################## -->
<!-- Definition of Environment Info Type     -->
```

```
<!-- ########################################### -->
<complexType name="EnvironmentInfoType">
  <sequence>
    <element name="GlobalPosition" type="arstv:GlobalPositionType"
minOccurs="0" maxOccurs="unbounded"/>
    <element name="Light" type="arstv:LightType" minOccurs="0"
maxOccurs="unbound"/>
    <element name="Camera" type="arstv:CameraType" minOccurs="0"
maxOccurs="unbounded"/>
  </sequence>
</complexType>
<!--################################### -->
<!--Definition of Global Position type  -->
<!--################################### -->
<complexType name="GlobalPositionType">
  <sequence>
    <element name="Address" type="mpeg7:PlaceType" minOccurs="0"/>
  </sequence>
  <attribute name="longitude" use="required">
    <simpleType>
      <restriction base="double">
        <minInclusive value="-180.0"/>
        <maxInclusive value="180.0"/>
      </restriction>
    </simpleType>
  </attribute>
  <attribute name="latitude" use="required">
    <simpleType>
      <restriction base="double">
        <minInclusive value="-90.0"/>
        <maxInclusive value="90.0"/>
      </restriction>
    </simpleType>
  </attribute>
</complexType>
<!--################################### -->
<!--Definition of Light type        -->
<!--################################### -->
<complexType name="LightType">
    <sequence>
       <element name="Position" type="ARSTV:DirectionType" minOccurs="0"/>
       <element name="Rotation" type="ARSTV:RotationType" minOccurs="0"/>
    </sequence>
    <attribute name="type" type="unsignedInt" use="optional"/>
    <attribute name="color" type="ARSTV:ColorType" use="optional"/>
    <attribute name="intensity" type="ARSTV:zeroToOneType"/>
</complexType>
<complexType name="PositionType">
    <attribute name="px" type="float" use="optional"/>
    <attribute name="py" type="float" use="optional"/>
    <attribute name="pz" type="float" use="optional"/>
</complexType>
<complexType name="RotationType">
    <attribute name="vx" type="float" use="optional"/>
    <attribute name="vy" type="float" use="optional"/>
    <attribute name="vz" type="float" use="optional"/>
</complexType>
<!--################################### -->
```

```
<!--Definition of Color type        -->
<!--################################### -->
<simpleType name="ColorType">
  <restriction base="NMTOKEN">
    <whiteSpace value="collapse"/>
    <pattern value="#[0-9A-Fa-f]{6}"/>
  </restriction>
</simpleType>
<!--################################### -->
<!--Definition of Camera type        -->
<!--################################### -->
<complexType name="CameraType">
  <attribute name="fov" type="float" use="optional"/>
</complexType>
```

I.10 User interaction

```
<!-- ######################################### -->
<!-- Definition of User Interaction Type     -->
<!-- ######################################### -->
<complexType name="UserInteractionType">
   <sequence>
      <element name="Resource" type="arstv:ResourceType" minOccurs="0"/>
     <element name="SRT" type="arstv:SRTType" minOccurs="0"/>
   </sequence>
   <attribute name="event" type="unsignedInt" use="optional"/>
</complexType>
```

Appendix II

Use case example of haptic device

(This appendix does not form an integral part of this Recommendation.)

With a haptic device, TV users can feel or enjoy the intensity/vibration of the content.

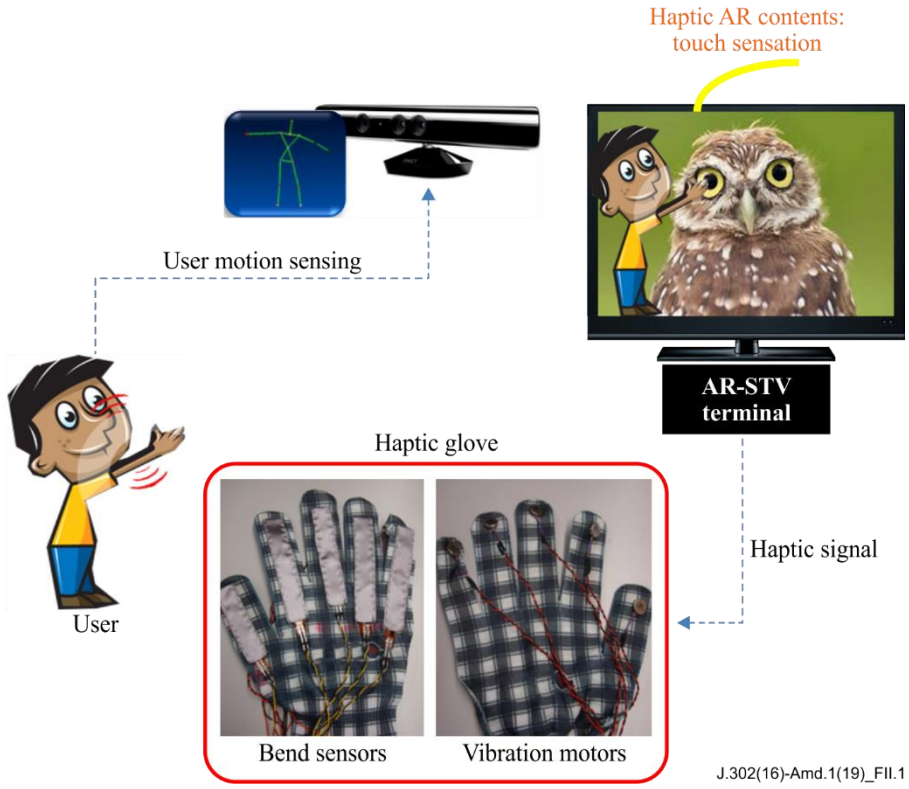


Figure II.1 – An example of AR-STV service connected with haptic device

AR-STV service can be extended beyond the TV screen to an external device such as a haptic device. On receiving a haptic signal from an AR-STV terminal, a TV viewer with haptic gloves, as shown in the case of Figure II.1, would have the tactile feel of 'feathers' by touching the AR content via user motion sensing.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	**Cable networks and transmission of television, sound programme and other multimedia signals**
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2019