3GPP TR 26.999 V0.3.0 (2020-01)

Technical Report

3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Virtual Reality (VR) Streaming Interoperability and Characterization (Release 17)

3GPP

Postal address

3GPP support office address
650 Route des Lucioles – Sophia Antipolis
Valbonne – France
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet
http://www.3gpp.org

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

In the present document, modal verbs have the following meanings:

shall indicates a mandatory requirement to do something

shall not indicates an interdiction (prohibition) to do something

The construc“ions ”shall“ and "sha”l not" are confined to the context of normative provisions, and do not appear in Technical Reports.

The construc“ions”"must“ and "mu”t not" are not used as substitute“ for ”shall“ and "sha”l not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

should indicates a recommendation to do something

should not indicates a recommendation not to do something

may indicates permission to do something

need not indicates permission not to do something

The constru“tion "m”y not" is ambiguous and is not used in normative elements. The unambiguous construc“ions "mig”t no“" or "sha”l not" are used instead, depending upon the meaning intended.

can indicates that something is possible

cannot indicates that something is impossible

The construc“ion” "can“ and "”annot" are not substitute“ fo” "may“ and "ne”d not".

will indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

will not indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

might indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

might not indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

is (or any other verb in the indicative mood) indicates a statement of fact

is not (or any other negative verb in the indicative mood) indicates a statement of fact

The construc“io”s "is“ and "”s not" do not indicate requiremes.

1 Scope

The present document provides content generation guidelines, performance and bit rate characteristics, as well as reference test material and test results for improved usability of [2] technologies based on a set of usage scenarios.

The specification [2] includes several VR media profiles for video and a single media profile for audio with different configuration options. The specification focuses primarily on interoperability requirements for VR360 applications, but does neither address content generation guidelines nor performance characterization of the solutions. In order for content providers and the rest of the ecosystem to be able to select and configure the technologies defined in [2] and to generate content for streaming applications, collecting such information would be most valuable. In addition, VR client implementers would benefit from conformance bit streams and media presentations that fulfil the requirements of the specification and are generated using the content generation guidelines. Bit streams and media presentations generated through these guidelines would be most useful.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21“905: "Vocabulary for 3GPP Specific”tions".

[2] 3GPP TS 26.118: “Virtual Reality (VR) profiles for streaming applications”.

[3] 3GPP TS 26.260: “Objective test methodologies for the evaluation of immersive audio systems”.

[x] <doctype> <#>[([up to and including]{yyyy[-mm]|V<a[.b[.c]]>}[onward“])]: "<”itle>".

3 Definitions of terms, symbols and abbreviations.

3.1 Terms

For the purposes of the present document, the terms given in 3GPP TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in 3GPP TR 21.905 [1].

<defined term>: *<definition>.*

example: text used to clarify abstract rules by applying them literally.

3.2 Symbols

For the purposes of the present document, the following symbols apply:

<symbol> <Explanation>

3.3 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in 3GPP TR 21.905 [1].

<ABBREVIATION> <Expansion>

4 Usage scenarios

4.1 Introduction

This clause defines a set of key usage scenarios for the application of technologies defined in [2] for live and on-demand streaming of VR360 content.

5 Content generation guidelines

5.1 Introduction

This clause documents content generation guidelines that address the usage scenarios for the different technologies and operating points from [2].

6 Source content material

6.1 Introduction

This clause documents relevant source content material for the usage scenarios.

6.2 Orange test sequence

6.2.1 The VR Experience

The VR 360 sequence is intended to be experienced through a VR headset. The VR spectator is immersed into the stage of a TV news. The real presenter welcomes the spectator and let him on his own 2 minutes before going live. The scenario has been defined to let the VR spectator feel the increasing pressure 2 minutes before the live broadcast as well as perceive the coordination of the technical team both on stage and in the control room in order to make such a well-known program possible.

Figure 1 illustrates the environment of the sequence.

Figure 1: Screenshot of the candidate test sequence

6.2.2 Video capture

Video was shot with a rig of 24 cameras (4 cameras to the top, 4 cameras at the bottom and 16 in a horizonal crown configuration). Each of the cameras had a 2.7K resolution with a 120° angle. Unlike the 8 cameras facing the top and the bottom, the 16 horizontal cameras worked in couples in order to create a stereoscopic effect. The recording was done on SD cards, thus generating 24 files to be synchronized altogether for each shot. Figure 2 below illustrates the camera rig configuration.

Figure 2: Video shooting configuration

6.2.3 Audio configuration

A 3D microphone made of 32 sensors was placed over the rig. It was linked via RJ45 to an interface delivering through FireWire towards a computer for recording. The audio source in HOA format allowed the recording of ambient sounds perfectly localized in 3D. A few lapel microphones were used as well as one audio signal to simulate the earpiece of the presenter to receive the control room information.

Figure 3 shows the audio capture system.

Figure 3: Audio capture configuration

6.2.4 Final product

The resulting version of the sequence is made of:

- Video
    1) 8k resolution
    2) 50 frames per second
    3) Equirectangular projection
    4) 8 bits per pixel, RGB
    5) BT.709 color space
    6) around 2min 10sec. duration
    7) Available in mono and stereo.

- Audio:
    1) French
    2) HOA 3rd order
    3) one stereo track for head lock.

6.3 InterDigital test sequence

6.3.1 The VR Experience

The VR 360 sequences are intended to be experienced through a VR headset. The experience is that of a biker or street walker strolling through a number of attractions and local community in San Diego, California, USA, including Gaslamp Quarter neighbourhood, the harbour, a park, an old trolley, and a local residential community. Figures 4 to 7 provide screenshots from the four full-8K sequences in raw video format.

Figure 4 Gaslamp360_8192x4096_30fps_300frames_8bits.yuv

Figure 5 Harbor360_8192x4096_30fps_300frames_8bits.yuv

Figure 6 Kiteflite360_8192x4096_30fps_300frames_8bits.yuv

Figure 7 Trolley360_8192x4096_30fps_300frames_8bits.yuv

Figure and Figure provide screenshots from the two 8K (7680x3840) sequences in compressed format.

Figure 8 Community_7680x3840_29.97fps_150mbps_5mins.mp4

Figure 9 Intersection_7680x3840_30fps_150mbps.mp4

6.3.2 Video Capture

The first four video sequences were captured using a GoPro Omni camera rig, a synchronized camera array with 6 GoPro Hero4 Black cameras, each camera can capture 2.7K resolution at 60fps. Kolor Autopano Video Pro was used to stitch the video captured by the GoPro camera rig. No encoding is done after stitching. Therefore, the stitched content is provided in uncompressed raw YUV format.

The last two video sequences were captured using Insta360™ Pro camera with 6 F2.4 fisheye lenses. The maximum 360 video capture resolution is 7680x3840 at 30fps with post-processing stitching. The MP4 bitstream generated by Insta360™ Pro camera were stitched and encoded by the camera, the compression rate is 150mbps with build-in HEVC encoder.

Figure and Figure illustrate GoPro Omni camera rig and Insta360™ Pro camera used to capture the aforementioned test sequences.

Figure 10 GoPro Video capture configuration

Figure 11 Insta360™ Pro Video Capture

6.3.3 Final Product

The resulting versions of the six sequences have the following characteristics:

- 8K resolutions (8192x4096 or 7680x3840)
- 30 frames or 29.97 frames per second
- Equirectangular projection
- 8 bits per pixel, YUV
- BT.709 color space
- from 10 seconds to 5 minutes

The MD5 checksum of each of the uncompressed 8-bit YUV420 sequences is listed below:

Sequence	**MD5**
Gaslamp360_8192x4096_30fps_300frames_8bits.yuv	858dfe4b7a2d463f1866c82dd14d51be
Harbor360_8192x4096_30fps_300frames_8bits.yuv	aa827fdd01a58d26904d1dbdbd91a105
KiteFlite360_8192x4096_30fps_300frames_8bits.yuv	18c0ea199b143a2952cf5433e8199248
Trolley360_8192x4096_30fps_300frames_8bits.yuv	84d6bfc93053ef28ddfcbe41d0864a9c

The proposed sequences are available for public download at https://www.interdigital.com/visual-technologies#

7 Objective quality measures and parameters settings

7.1 Introduction

This clause defines objective quality measures for the relevant usage scenarios and provides information of the technologies in [2] and certain parameter settings such as bitrates, quality options, etc. for each media type.

8 Conformance

8.1 Introduction

This clause provides bit streams and media presentations that follow the content generation guidelines and conform to the requirements in [2]. Also, it documents quality, required bitrates and QoS and other characteristics of the conformant bitstreams.

9 Test results

9.1 Introduction

This clause documents test results for the relevant usage scenarios, and characterizes the applicability of the technologies from [2] and certain parameter settings such as bitrates, quality options, etc. for each media. Also, based on the usage scenarios, this clause defines a test and characterization framework for the technologies in [2] building on the initial experience developed in [3] (for audio). This clause includes also subjective test results for the relevant usage scenarios and characterize the applicability of the technologies in [2] and certain parameter settings such as bitrates, quality options, etc. for each media.

4 Examples for styles

Abc

4.1 Heading styles

Heading styles are included in the 3GPP TS Template and are used as follows:

Do not use any built-in automatic numbering *for 3GPP documents. Although this is sometimes useful in the early drafting stages of a document, once the document has been placed under change control, the clause numbering needs to be fixed in order to keep cross-reference consistency as the 3GPP specification set develops.*

Heading 1: Used for Main clauses (1, 2, 3, etc.). Also used for Annex clauses (A.1, A.2, etc.).

Heading 2: Used for Main clauses (4.1, 4.2, 5.1, 5.2, etc.). Also used for Annex clauses (A.1.1, A.1.2, etc.).

Heading 3: Used [f]or 2nd level clauses (4.1.1, 4.1.2, 5.1.1, 5.1.2, etc.). Also used for Annex clauses (A.2.1.1, A.2.1.2, etc.).

Heading 4 & 5: Used [f]or 3rd [a]nd 4th level clauses and Annex clauses.

Heading 6 & 7: Not used, instead use “ty”e "H6" so that the title appears in the document, but does not appear in the Table of Contents.

Heading 8: Used for Main Annex titles in Specifications (3GPP TS) (e.g. Annex A (normative):).

Heading 9: Used for Main Annex titles in Reports (3GPP TR) (e.g. Annex A:).

4.2 Other common styles

Normal: Used for main document text.

NO: Used for Notes in the text (Allows Tab and Indent). See example below.

NW: Same as NO, but Without line space after. Used when there are many notes in sequence.

NOTE 1: This is an example of a note formatted in style NW. The style is designed to allow space for note numbering and line wrap with a hanging indent. There is no line space after.
NOTE 2: This is an example of a note formatted in style NO. The style is designed to allow space for note numbering and line wrap with a hanging indent. There is a line space after.

Bullet styles: The following bullet styles are provided.

B1: Bullet level 1 for main bullet points.

B2: Bullet level 2 for sub bullets.

B3-B5: for further sub bullets.

NOTE: Bullets are usually formatted manually, using a hyphen (-) or alphanumeric identifiers: a), b), or 1), 2) etc. followed by a tab character. **Automatic bullet features should not be used** as they may be lost if template styles are re-applied later.

Table styles: **TAH**, **TAL**, **TAC**, **TAR**, **TAN**, for **TA**ble **H**eaders, **L**eft justified, **C**entred, **R**ight justified and **N**otes in tables: Style **TH** is used for the **T**able **H**eading (title or caption). See example below.

Table 1: Example of table styles

Col 1 Header (TAH)	Col 2 Header (TAH)	Col 3 Header (TAH)
Left Justified (TAL)	Centred (TAC)	Right Justified (TAR)
NOTE: A special style is provided for notes within a table (TAN).		

Warning: The default setting for table cells is to disallow rows to break at a page boundary. If you include tables with very long cells, likely to extend beyond the bottom of the page (bearing in mind the table header and the page header and footers, and the margin settings), then you must enable th't "ow's "Allow row to break across"pages" setting.

Figure styles: Figures and graphics are formatted with "ty"e "**TH**" which keeps the figure with the following paragraph, usually the figure title. **F**igure **T**itles (captions) are formatted with "ty"e "**TF**". See example below.

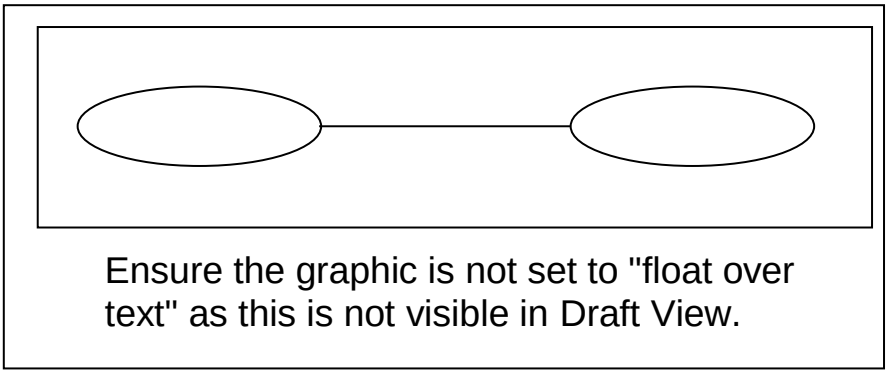

Figure 1: Example figure layout. To r"move "float ove" text" select the graphi" and "Format Obje"t–..." - De-s"lect "float ove" text" in the Position Tab

Annex A:
Process steps for video

Required Processes	Status Check and Proposal
Test Sequences with the following parameters: - Basic Video Profile o 4096 × 2048 o BT.709 o 50Hz, 60Hz o 8 bit, 4:2:0 o May be processed to meet lower requirements - Main Video Profile o Mono: 6144 × 3072 o Stereo: 3840 × 1920 o BT.2020, BT.709 o 50Hz, 60Hz o 10 bit, 4:2:0 o May be processed to meet lower requirements - Flexible Video Profile o Mono: 8192 × 4096 o Stereo: 4320x2880 o BT.2020, BT.709 o SDR, HDR o 50Hz, 60Hz, 100Hz, 120Hz o 10 bit, 4:2:0 o May be processed to meet lower requirements Other Requirements: - At least 10 seconds duration - Encoded bitstreams can be published as part of a 3GPP TR	Status Check Interdigital informs on these sequences: o Gaslamp360_8192x4096_30fps_300frames_8bits.yuv o Full 360-degree ERP raw video sequence, resolution 8192x4096, frame rate 30fps, 4:2:0 format, bit depth 8, duration 300 frames, SDR, color space BT.709, no audio o Harbor360_8192x4096_30fps_300frames_8bits.yuv o Full 360-degree ERP raw video sequence, resolution 8192x4096, frame rate 30fps, 4:2:0 format, bit depth 8, duration 300 frames, SDR, color space BT.709, no audio o Kiteflite360_8192x4096_30fps_300frames_8bits.yuv o Full 360-degree ERP raw video sequence, resolution 8192x4096, frame rate 30fps, 4:2:0 format, bit depth 8, duration 300 frames, SDR, color space BT.709, no audio o Trolley360_8192x4096_30fps_300frames_8bits.yuv o Full 360-degree ERP raw video sequence, resolution 8192x4096, frame rate 30fps, 4:2:0 format, bit depth 8, duration 300 frames, SDR, color space BT.709, no audio o Balboa360_6144x3072_60fps_600frames_8bits.yuv o Full 360-degree ERP raw video sequence, resolution 6144x3072, frame rate 60fps, 4:2:0 format, bit depth 8, duration 600 frames, SDR, color space BT.709, no audio o Broadway360_6144x3072_60fps_600frames_8bits.yuv o Full 360-degree ERP raw video sequence, resolution 6144x3072, frame rate 60fps, 4:2:0 format, bit depth 8, duration 600 frames, SDR, color space BT.709, no audio o Community_7680x3840_29.97fps_150mbps_5mins.mp4 o Full 360-degree ERP HEVC video bitstream (150mbps), resolution 7680x3840, frame rate 30fps, 4:2:0 format, bit depth 8, duration 5 mins, SDR, color space BT.709, no audio o Intersection_7680x3840_30fps_150mbps.mp4 o Full 360-degree ERP HEVC video bitstream (150mbps), resolution 7680x3840, frame rate 30fps, 4:2:0 format, bit depth 8, duration 5 mins, SDR, color space BT.709, no audio o All sequences are available at https://www.interdigital.com/video-resources/, and can be further downsampled to 4K resolution or other projection format using JVET 360Lib software.
Content Generation Guidelines	Status check • A few statements in [2].

	Proposal: • Document content generation guidelines that can be used for the development of test material
Test Case Definition for running subjective tests - Needs to be a well-defined subset only - Should be based on the same content, so 8k is needed. - This is for characterization, not for selection. Hence, comparable and accessible tools should be used such as reference software of MPEG, etc.	Status Check: o Nothing available until now Proposal: o Same content prepared for each media profile with the following parameters: o Simple profile, viewport-independent o Main profile, viewport-independent o Main profile, viewport-dependent ▪ Different viewport-switching latencies o Advanced profile, viewport-dependent ▪ One or two different configurations o Different bitrates, but no dynamic bitrates. o Mono, stereo o All cases must be supported by o Content generation guidelines o conformance bitstreams
Test Material Preparation o Encoding and decoding at different bitrates o Inclusion of relevant metadata o Exact definition of encoding parameters, preferably with a reference software	Status: o Nokia provided the OMAF Creator SW. More information is available in the TR 26.928. Proposal: o Analyze the above SW o Check further work in MPEG VVC and HEVC o Document exactly how the material was prepared for tests.
Subjective Test Run Definition: o Clear definition of subjective tests o Emulate the testing by providing the decoded sequences on high-power PC	Status Check: o We have tests on [2], clause 7.2 Proposal: o Neglect for now. Focus on interop and objective measures. o Should be based on already executed tests o Do not include this as part of the study item as necessity, but make it nice to have.
Test Run Execution: o Preferably done in a fair manner o No selection, so less critical	Proposal: o not include this as part of the study item as necessity, but make it nice to have.

o Preferably done by multiple parties	
Objective Measures o Define reasonably good objective measures o Don't claim that they match subjective quality, but permit some amount of comparison	Proposal: o Investigate what MPEG has done for HEVC and VVC o Document the measures o Apply these measures to the subjective tests to identify quality. o Do a second set of tests for flatscreen rendering that applies objective PSNR measures.
Test Case Definition for conformance bitstreams - should be much richer and cover different aspects - including DASH Preparation - coverage restrictions, etc.	Proposal: o include conformance bitstreams for all tests done above o make this lower priority as it not essential, but permit the documentation
Hosting of Material o Expect huge amount of data o Some information needs to only be maintained temporarily o Conformance streams should be available permanently	Proposal: o Learn from MPEG and DASH-IF and apply the appropriate procedures o Ask DASH-IF to host test material

Annex B:
Test Vectors

B.1 Introduction

In the context of this Report, test vectors are provided. As these test vectors are of significant size, they are hosted by DASH-IF (http://www.dashif.org). DASH-IF is always interested to support the industry in interoperability efforts on DASH-related matters. However, note that DASH-IF is not able to provide any service and availability guarantees of such vectors. Finally, test vectors are preferably be provided following the licensing terms of DASH-IF Test assets.

In order to add test vectors, clause B.2 provides some procedures on how to host test vectors on DASH-IF website.

B.2 Uploading and Hosting Test Vectors

The following information is provided by the DASH-IF Test Asset Coordinator. For information on how to contact the test asset coordinator, please communicate with the co-rapporteur of the study item.

1. How hosting/uploading can be done?
    1. The preferred way is that the hosting is done on the CDN server by the DASH-IF Test Asset Coordinator.
    2. Uploading to the test assets database would be done:
        1. Preferably by 3GPP contributors
        2. Or alternatively by the DASH-IF Test Asset Coordinator.
    3. Maintenance would preferably be done by 3GPP Contributors via DASH-IF Test Asset Coordinator assigning the contributors to the github issue or us sending them an e-mail.
2. What information is needed
    1. For hosting, DASH-IF Test Asset Coordinator would need the related MPD files and the media content that each MPD points to be sent to them. They would also need the space required by the content (MPD + media).
    2. For uploading:
        1. Meaningful categorization should be done.
            1. Feature Group (3GPP-VR_CoGui), Feature, Test Case and Test Vector field naming should be meaningful.
        2. Preferably by 3GPP Contributors: DASH-IF Test Asset Coordinator would provide 3GPP Contributors with the MPD URLs on the Akamai server. DASH-IF Test Asset Coordinator would create 3GPP Contributors an account on the test assets database. 3GPP would add the test vectors by following the 1st bullet on categorization.
        3. If by DASH-IF Test Asset Coordinator: 3GPP Contributors provide DASH-IF Test Asset Coordinator with the categorization in an excel sheet by following the 1st bullet. DASH-IF Test Asset Coordinator add the test vectors.
    3. For maintenance, DASH-IF Test Asset Coordinator needs to gather the issues with the test vectors and point 3GPP Contributors to the issue via e-mail or assign them the issue if they have a github account. For this, DASH-IF Test Asset Coordinator would need an e-mail address or a github user account name.
3. What is the license that is commonly used

1. Creative Commons Attribution - NonCommercial-NoDerivatives 4.0 International license

4. Any other information on the test data base.

    1. None

Annex C (informative): Change history

Change history							
Date	Meeting	TDoc	CR	Rev	Cat	Subject/Comment	New version
2019-10	SA4#106	S4-191285				First version. Added agreed content from S4-191139 and S4-191287	0.1.0
2019-10	SA4#106	S4-191294				Updates to section 6.2 (Orange test sequence)	0.2.0
2020-01	SA4#107	S4-200206				Added section 6.3 (InterDigital test sequence)	0.3.0
2020-01	SA4#107	S4-200208				Added Annex B on Test Vectors	0.3.0