ETSI TR 126 929 V17.0.0 (2022-05)

TECHNICAL REPORT

5G;
QoE parameters and metrics relevant to the Virtual Reality (VR) user experience
(3GPP TR 26.929 version 17.0.0 Release 17)

Reference

RTR/TSGS-0426929vh00

Keywords

5G

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Legal Notice

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

The present document may refer to technical specifications or reports using their 3GPP identities. These shall be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between 3GPP and ETSI identities can be found under http://webapp.etsi.org/key/queryform.asp.

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

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.

Introduction

User experience based network management is important for operators so that they can provide best quality services. In case of a VR service, which is much more complex than traditional video streaming in terms of content creation, network transmission and device requirements, the provision of a satisfying immersive experience is more challenging. The first step of this effort would be to find out what factors have an impact on user experience, and define reference metrics and parameters that would help operators make assessment of user experience, do trouble shooting and design target solutions.

The analysis of impact on user experience involves the whole E2E VR service chain:

- Creation of content. VR content creation would involve multiple steps such as capture, stitching, projection, and encoding, each step would have an impact on the content itself. It is necessary to look into each step and find out what factors are relevant to the VR experience;
- Network transmission. The amount of video data of VR content entails a high streaming bitrate, and may lead to a risk of network and/or access link congestion and re-buffering, and thus an impediment to limiting latency. Latency is one of the key elements to create the feeling of immersiveness, and thus has considerable impact on user experience.
- Device requirement. Compared with a traditional device, be it a mobile phone or a tablet, a VR device exhibits many more attributes, designed to help create immersive experience. The degree of freedom it provides to users, the sensitivity of sensors that enable quick catching of head movement, and many other attributes, all need to be studied and evaluated about their relevance to user experience.

The present document investigates the QoE metrics relevant with VR experience from the aforementioned three aspects, and also the way of reporting these QoE metrics to the network for further analysis.

1 Scope

The present document provides a study on the QoE metrics relevant to VR service. The study focuses on:

- Defining a device reference model for VR QoE measurement points.
- Studying key performance indicators that may impact the experience of VR service.
- Identifying the existing QoE parameters and metrics defined in SA4 standards such as TS 26.247, TS 26.114 which are relevant to Virtual Reality user experience;
- Identifying and defining new QoE parameters and metrics relevant to Virtual Reality user experience, taking into consideration the use cases listed in TR 26.918, and any sources that show the relevance of new metrics, e.g. scientific literature, specifications/solutions from other standard organizations.
- Analysing potential improvements to the existing QoE reporting so as to better accommodate VR services.
- Providing recommendations to future standards work in SA4 on the QoE parameters and metrics and, as necessary, coordinate with other 3GPP groups and external SDOs, e.g. MPEG, ITU-T.

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 Specifications".

[2] 3GPP TR 26.918: "Virtual Reality (VR) media services over 3GPP".

[3] 3GPP TS 26.247: "Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH)".

[4] ISO/IEC 23009-1: 2014/Amd. 1:2015/Cor.1:2015: "Information technology -- Dynamic adaptive streaming over HTTP (DASH) -- Part 1: Media presentation description and segment formats".

[5] Recommendation ITU-T P.1203 (10/2017): "Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport".

[6] A. Singla, S. Fremerey, W. Robitza, A. Raake, "Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays", in 9th International Conference on Quality of Multimedia Experience (QoMEX), May 2017.

[7] A. Singla, A. Raake, W. Robitza, P. List, B. Feiten, "AhG8: Subjective Quality Evaluation for Omnidirectional (360°) Videos", Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-G0152, 7th Meeting, July 2017.

[8] Recommendation ITU-T P.910 : "Subjective video quality assessment methods for multimedia applications", International Telecommunication Union, April 2008.

[9] Recommendation ITU-R BT.500-13: "Methodology for the subjective assessment of the quality of television pictures", International Telecommunication Union, Jan. 2012.

[10] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," The International Journal of Aviation Psychology, 1993.

[11] E. Asbun, Y. He, Y. He, "AHG8: InterDigital Test Sequences for Virtual Reality Video Coding", Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0039, 4th Meeting, Oct. 2016.

[12] S. Schwarz et. al., "Tampere pole vaulting sequence for virtual reality video coding", Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0143, 4th Meeting, Oct. 2016.

[13] W. Sun, R. Guo: "Test Sequences for Virtual Reality Video Coding from LetinVR", Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0179, 4th Meeting, Oct. 2016.

[14] 3GPP TS 26.118: "3GPP Virtual reality profiles for streaming applications".

[15] ISO/IEC 23090-6: "Immersive media metrics".

[16] 3GPP TS 26.114: " IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction".

[17] 3GPP TR 26.909: "Study on improved streaming Quality of Experience (QoE) reporting in 3GPP services and networks".

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].

3.2 Symbols

Void.

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].

AV	Audio/video
AVC	Advanced Video Coding
3DOF	3 Degrees of freedom
6DOF	6 Degrees of freedom
DASH	Dynamic Adaptive Streaming over HTTP
FHD	Full High Definition
FOV	Field of view
HMD	Head Mounted Display
MCC	Metrics collection and computation
MPD	Media Presentation Description
QoE	Quality of Experience
VR	Virtual Reality
UL	Up-link

4 VR QoE overview

4.1 Introduction

The VR service is much more complex than traditional 2D video streaming, and thus defining relevant metrics is also more challenging. The following clauses address some of these challenges.

4.2 VR QoE metrics evolution

The goal with this study is to suggest improvements to the existing QoE reporting, so that suitable QoE metrics are available to better understand the VR service quality as experienced by the VR users. A complicating factor for the study is the lack of a thorough scientific understanding on exact how different VR conditions and impairments relate to the final user quality.

As VR services becomes more mature, and also more standardized, this understanding will become more clear over time. It is not unlikely that in a long-term perspective there might be standardized objective quality models, similar to the ITU-T P.1203 [5], which translate measurable QoE metrics into the final user experience.

However, there is a significant delay from the time a standard (3GPP, MPEG, ITU-T etc.) is ready, until the corresponding standard features are actually implemented large-scale by the device industry. There is also a further delay until the penetration of such standard-compliant devices gets high enough to become useful for wider analysis.

Thus, waiting with defining VR-related QoE metrics until a full and complete understanding has been gained will cause a significant gap in time, where VR service-quality monitoring cannot be adequately performed, at least not by standardized means.

The question is how to move forward during this transition period, where a detailed understanding of the QoE relations is not always available? And how to do it in a way where stepwise QoE metrics refinement can be done over several releases?

The key for how to succeed with this can actually be found in the existing QoE metrics in TS 26.247. When these metrics were standardized in Rel-10, several of the defined "metrics" are actually not really stand-alone QoE metrics. Rather they are just lists of events initiated by the client or by the user.

A typical example is the PlayList, which contains user and client actions, together with timestamps and a minimal set of related metadata. As discussed in earlier SA4 meetings, these "event lists" are not QoE metrics by themselves, and there has also been suggestions to enhance TS 26.247 with additional "real QoE metrics".

However, a big advantage with the event lists is that they contain the basic information necessary to calculate other derived QoE metrics. For instance, ITU-T P.1203 needs as one input a metric containing the number of rebufferings. As P.1203 was not fully standardized until 2017, there was no knowledge during the Rel-10 QoE work that such a metric would later be needed. However, due to the event-based PlayList it is now actually possible to derive this metric, without changing the TS 26.247 standard. The same is true for several other metrics needed by P.1203.

A drawback with event lists, especially for the VR case, is that they can potentially contain a lot of events. For normal 2D streaming the interaction from the user is much more limited, basically play, stop, jump, etc. For a VR service all of those are also there, but the main interaction is continuous head (or even body) movements, which then might also cause the player to interact towards the network in different ways (fetching different tiles etc.).

However, there are several ways to handle the report size problem, for instance using a constant (but configurable) sample time for the "movement list", and/or a (configurable) movement threshold before a movement is logged. Although it might be difficult to define exactly what sample time or movement threshold that would be the best compromise between accuracy and report size, actually there is no need to know that right now. It is a decision which can be made (years) later by the operator or the service provider, depending on the QoE use case at that point in time.

When VR services become more mature, it is expected that more optimized versions of the QoE metrics can possibly be standardized, resulting in more tailored and smaller metrics. But having a basic QoE reporting available as early as possible is a big advantage, as it is typically during the early phases of a new service adoption that the need for quality-related feedback and analysis is most important.

Thus a stepwise evolution of VR QoE metric could be to use event lists whenever possible for the initially defined metrics, as this seems to be the most future-proof representation. The lists could allow some basic configurability to enable a flexibility between accuracy vs. report size. This would allow later derivation of more specific QoE metrics, as well as other quality-related aspects, some of which might not even be known today.

4.3 Viewport-related QoE aspects

4.3.1 Introduction

From the user point of view, one of the main differences between normal 2D video streaming and VR video streaming is the notion of a viewport. Instead of always seeing the complete video, the user only sees a cropped part of the video, the viewport, depending on the direction of the device.

The resulting impairments and quality experience will vary depending on content authoring strategy, the client rendering strategy, and any network impact. The following clauses show some (non-exhaustive and simplified) examples of possible delivery scenarios and resulting impairment types. Note that while the viewport also affects audio, no audio aspects are included in the examples.

The examples should not be seen as any endorsement of specific authoring or delivery technology, they only illustrate some possible impairments as seen from the user point of view. Also, for the sake of simplicity, all examples are drawn with square regions, but in practice regions can have different shapes and also do not need to have clear quality boundaries. Any spatial distortion due to the mapping into spherical rendering is also not considered.

4.3.2 Single stream region-independent coding

In this scenario the content is encoded with the same resolution and quality settings over the complete 360 content, and delivered as a single stream. The client decodes the complete stream but shows only the cropped part corresponding to the viewport (the red rectangle).

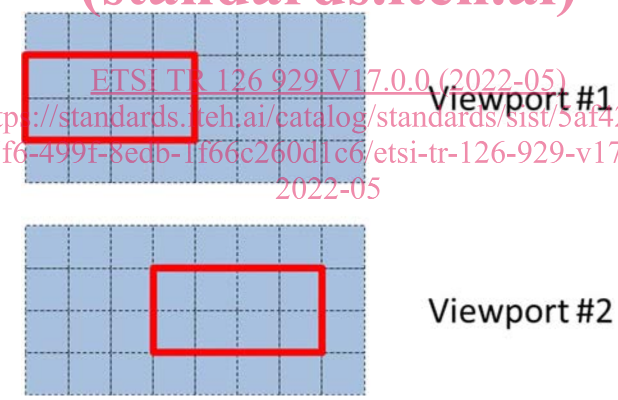

Figure 4.3.2: Single stream region-independent coding

In the example above the user moves the viewport from left to right, but as the encoding is the same for any viewport, this scenario is very similar to the 2D case. The main additional impairments would likely be related to projection artefacts and any device-internal rendering delay.

4.3.3 Single stream region-dependent encoding

The content is encoded with emphasis on one or more regions, where the content producer believes that most users will direct their viewport. Thus the resolution and/or coding quality is higher for selected parts of the spatial area, corresponding to these regions. The video is still delivered as one single 360 stream, and the client decodes and shows a cropped part corresponding to the viewport.

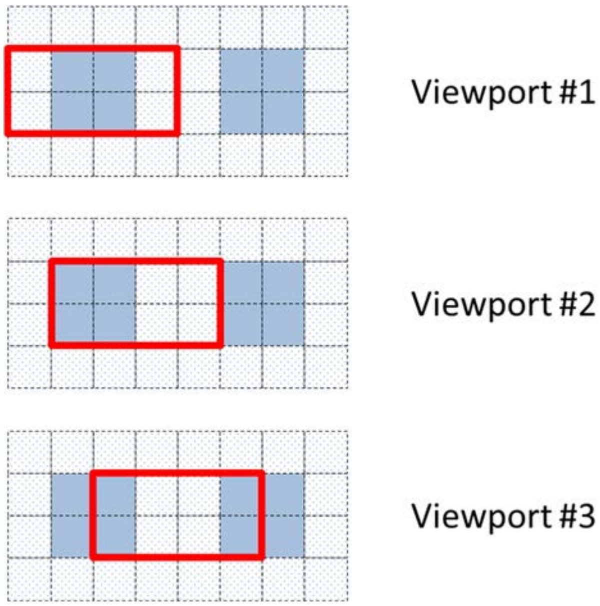

Figure 4.3.3: Single stream region-dependent encoding

In the example above, two emphasized regions are defined (dark grey), and the user moves the viewport from left to right. For all three viewports the average viewport quality is the same (i.e. 50 % high quality and 50 % low quality), but it is likely that the quality in the central part of the viewport has largest importance for the user. Thus you would expect viewport #1 to be experienced as best, and viewport #3 as worst.

4.3.4 Multiple stream region-dependent encoding

Multiple streams can be used, each emphasizing a given region. The receiver selects to download and render the stream which emphasized region best corresponds to the actual viewport.

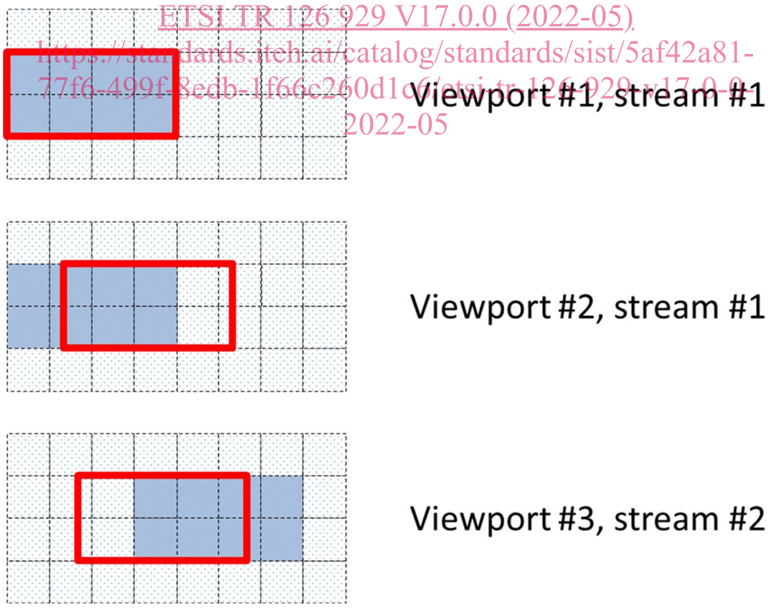

Figure 4.3.4: Multiple stream region-dependent encoding

In the example above, stream #1 and stream #2 have 25 % overlapping region-optimized encoding, and when reaching viewport #3 the receiver decides to switch to stream #2. Thus while turning the head between viewport #2 and #3, the user will see an instant change of quality for the left part (going high to low) and right part (going low to high) of the viewport.

Note that the average viewport quality is the same for viewport #2 and #3 (i.e. ~67 % high quality and ~33 % low quality) but the dynamic effect of switching between them is probably clearly visible. If the user moves his head back and forth between viewport #2 and #3, and the receiver selects to switch streams, such quality changes can likely be rather annoying.

4.3.5 Region-based encoding, simple head movement

With region-based encoding the client typically fetches viewport regions with high quality, while using low-quality regions for the background around the viewport (or even for the complete 360). The figures below (left and right) illustrate two variants of a simple head movement.

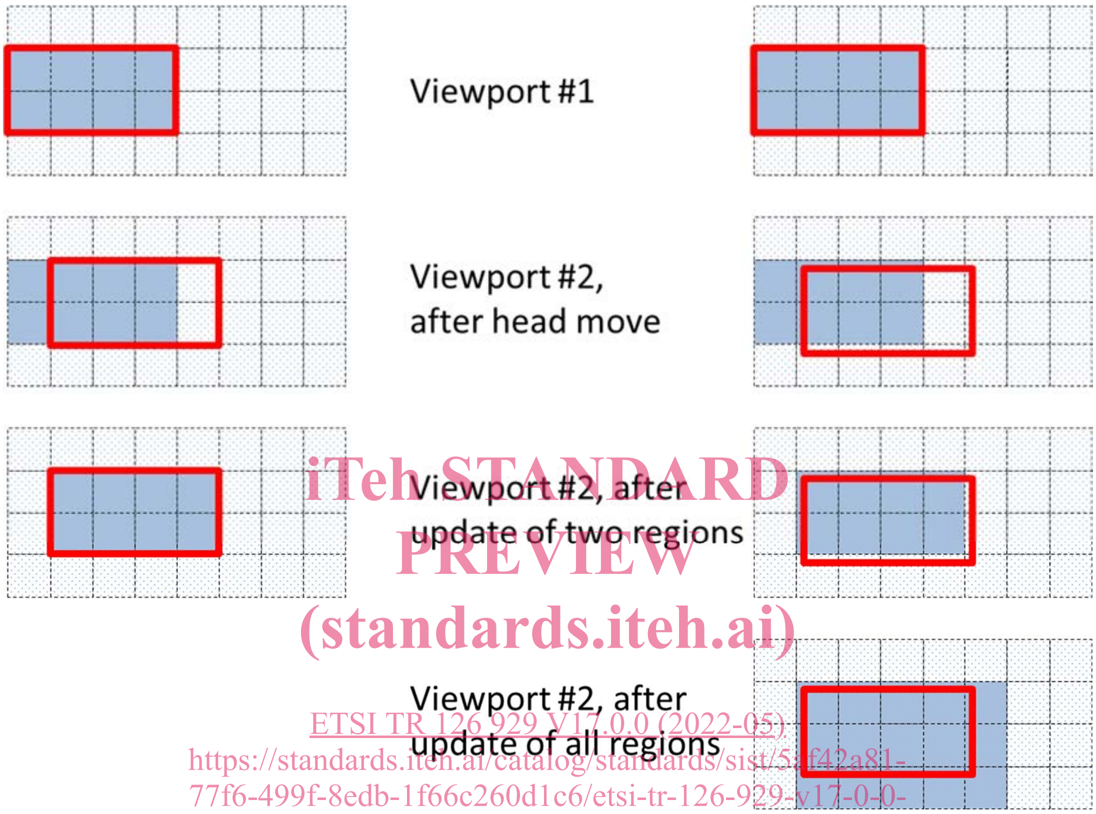

Figure 4.3.5-1: Region-based encoding, simple head movement

On the left side, the head movement is aligned with the regions, while on the right side it is moved a small bit into the next region column, and also a bit downward. In both cases the main impairment is the visible delay before high-quality versions of the new viewport regions have been fetched and rendered.

In the example it likely takes a bit longer to update the right scenario (nine regions instead of two), but due to the minimal viewport coverage of the seven outer regions, the user is unlikely to note any quality difference between the left and right scenario after the update of the first two regions. Thus although the final update delay probably is different, the experienced quality might be the same.

Note that the client could in principle instead decide to skip updating the seven outer regions due to their minimal viewport coverage. Alternatively, the client could decide to update them, but use an intermediate quality level instead of the highest quality, resulting in the example below.

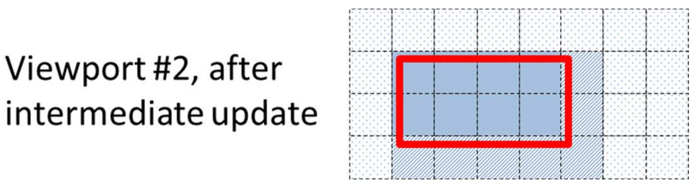

Figure 4.3.5-2: Region-based encoding, simple head movement, use of an intermediate quality level