Author: Dr.-Ing. Falk Schiffner - DLR-KOB **v_2.4 2021-10-20**

WORKING DOCUMENT – DRAFT

VIDEO QUALITY ASSESSMENT AND ESTIMATION FOR SPACE VIDEO

1. Introduction

This document is a meant as a starting point for a deeper discussion in CCSDS SIS-MIA working group on the topic of video quality.
As discussed in the last online meeting, there is a demand for more details on video quality requirements for space video, to support future iterations of the CCSDS recommendations and reports and/or to act as the source of a future green/orange book.

In this document, some possibilities and obstacles are discussed which must be considered when talking about introducing models, metrics, and estimation techniques for video quality assessment for space operations.
As known from the current recommendations and informational reports, video transmission in the domain for space flight can be subdivided in several use-cases [1]. These use-cases have partially completely different technological setups and requirements. This makes it tremendous difficult, if not impossible, to develop a quality model or metric that is applicable for all use-cases.

2. Definition of Quality, QoS and QoE

Because of the ever-growing importance of communication services, providers and technology developers are becoming more and more concerned about their products' quality. The main question which arises in quality analysis is: "how the product and its usage is perceived by the user group." Therefore, in order to provide the "best" quality, it is necessary to determine what the user considers "good quality" and measure the products and services accordingly.

In order to go deeper into the matter, terminology needs to be clarified first. The term quality is defined in [2] as:

> "Quality is a result of a comparison between the perceived composition of an entity and the desired composition."

Quality is only measurable with the help of people who are able to perceive and judge the content. Even instrumental quality estimators are based on human assessment as a ground truth. The assessment process is highly complex, and it depends on the perception and assessment situation. Accordingly, human factors such as expectations, experience, mood, and knowledge all play a role. The entity does not necessarily have to be a physical object. It can also be a service (such as a telephone service), which has features for which the quality may be quantified. Regarding the literature, the terminology is often unclear. The term Quality of Experience (QoE) is used when in contrast, Quality of Service (QoS) is more appropriate. The truth is, that the terms are not clearly separated, and QoE [3] is often used due to marketing reasons. Nevertheless, it is important to distinguish them. The term QoS defined in ITU Rec. E.800 [4] as:

> "The totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service."

ETSI's definition of QoS is quite similar to the definition from ITU. The IETF has an even a stronger network-centric definition on QoS, where the user is not mentioned at all:

> "A set of service requirements to be met by the network while transporting a flow."

On the other hand, the term QoE is defined in ITU-T Rec. P.10 [5] as:

> "The overall acceptability of an application or service, as perceived subjectively by the end-user."

Another similar definition comes from ETSI (TS 103 294) [6]:

> "Degree of delight or annoyance of the user of an application or service."

QoS is often used in practice when describing concepts and measures for network performance (like jitter, delay). In some instances, it is also used regarding forms of traffic engineering, such as rate policing. QoE, as a term, is often misused in ways which no longer fit the original definition. In the literature, one can read about "improving the QoE," where, for example, a technology for reducing the delay is reported. This could indeed lead to a better QoE but is still more related to QoS because it does not take the user, the context, etc., into account. If the relationship between QoS and QoE is understood correctly, they can enhance the control of network performance. Service providers could benefit when shifting from provider-centric QoS to user-centric QoE because it gives a more profound insight into quality influencing factors beyond the sheer technical scope.

Summarizing, both QoS and QoE reflect different viewpoints on the quantification of quality, although intersections exist between both. As a rule of thumb, QoE is referring the user's point of view, whereas QoS reflects the network/service provider's point of view. This should be kept in mind when talking about quality in general.

3. A brief summary of approaches for "Instrumental quality estimation"

This section is partially taken from [7].

An approach to estimate the perceived quality of a transmitted media system is the use of quality models. Often these models are also referred to as *objective models* or *instrumental estimation*. These models use signal measurement and estimation to predict the quality as rated by a person. There are several different approaches to do so, depending on the targeted use case. Here, an overview of the central concepts of instrumental quality estimation is given.

3.1 Video Quality Estimation

The different models can be categorized: First, the models can be grouped depending on the amount of information they need from the source signal.

1. Full-Reference: In this method, the reference signal is required as well as the transmitted signal. Here the two signals are compared, and with the help of quality metrics, the differences are quantified. If the difference is significant due to impairments, the resulting estimation of the quality rating will be low.

2. Reduced-Reference: Here, only some low-level features of the source signal are extracted. This information is then compared with corresponding low-level signal features of the transmitted signal. Several different proposed models use different numbers of features and extracting methods to estimate a quality score. When compared to full-reference model, a reduced amount of data is necessary.

3. No-Reference: As the name suggests, the source signal is not needed to estimate the quality. In other terms, no comparison between source and the transmitted signal is performed. The quality ratings are purely done by analyzing the transmitted signal. This bears the danger that actual content can be regarded as an impairment and can reduce the quality score (e. g. a chessboard is interpreted as block artifact).

This leads to the second way to group these models. Here, the classification is done by the type of input the models need.

1. Planning models: As the name suggests, these types of models are used in the planning process, with the consequence that only information about the used technologies (e.g. network configuration, planned bitrate, coding scheme) is available, and no actual measurements.

2. Packet-header-based models: Here, information is taken from the packet headers. These are extracted from the stream and often used in monitoring processes.

3. Bitstream-based models: These models need the bitstream fully or only partially decoded.

4. Signal-based models: Here, the decoded signal is required. These models extract information from the signal e. g. the pixel information of a frame or frequency range of the audio.

There are also combined hybrid models, that e. g. combine signal-based models and bitstream-based models. Other combinations are also conceivable and in use. Moreover, quality models can be categorized depending on the application they are targeted for (e. g. streaming service, service monitoring).

Model Development

In the development process of a quality model, the model type is selected dependent on the target. The creation can be roughly divided into three phases: First, the selection of the variables and parameters is taken into account; second, the construction of the model itself; and third, the evaluation of the model. There should be no arguing that the test set for the evaluation should reflect the scope of the model (e.g. the scenario the model should reflect). Each of the mentioned phases can then further split into details, like the choice of the modeling method, and it depends on the needs in each case.

Feature-based Approach to Modeling

This approach is often used by full-reference models and uses features from the source signal and the transmitted signal. Typically, the two signals are compared e. g. frame by frame or pixel by pixel. The video-features could be the luminance, color value, etc. For the set of features that are extracted, combined, or grouped, metrics are calculated to interpret the sequence under test.

Degradation-based Approach to Modeling

The idea behind that approach lies in finding a metric or a set of metrics for each perceptual quality dimension. Afterward, combining all perceptual quality dimensions estimates the overall quality. Degradations are, in most cases, regarded as components which reduce the quality. In this strategy, variables such as the network/transmission parameters under test, are used as input to impair the quality. The variables selected are dependent on the target application and the relevance and contribution of affecting the perceptual dimension must be considered. From this, the impact on quality is generalized for the type of input or similar input.

Subjective results-based Modeling

Here, the modeling approach is to map selected variables to subjective quality ratings. This strategy is used in parametric-based quality models. The mapping is typically done by using a regression analysis where the calculated values are aimed to match the subjective ratings. The resulting mapping function could be linear or polynomial. The goodness of fit depends on the chosen variables. The available number variables are limited to the number of parameters. The regression is often done in terms of "degradation-based modeling strategy" using the quality contribution terms related to different types of variables.

Model development impacting factors have been presented. The choice of the "right" approach and the "right" variables is often an iterative process. The main goal is always to predict the quality as close as possible, matching the perceived quality gathered from the test participants in the underlying experiments.

4. Video Quality for Space

As indicated in the last sections, making statements about the video quality is not a trivial task.
There are several aspects to consider, before choosing a modelling approach or quality metric.
A meaningful result can be obtained only if the boundary conditions are defined clearly in the beginning.

- Use-case for the modeling (subsystems vs. overall system):
    - Planning model
    - Monitoring model
    - . . .

- Available data:
    - Availability of the signal itself
    - Availability of only meta data of the signal
    - Availability of only transmission parameters
    - Availability of the reference signal?
    - . . .

As we know that there are six main scenarios for space video as stated in the CCSDS Recommendation [1]:

- Personal Conferencing
- Medical conferencing
- Proximity operations / situational awareness
- Public affairs
- High resolution imaging
- Crew training and instructions

It is necessary to investigate each scenario separately for possibilities to measure, assess, and estimate video quality regarding the use case and the available data.

5. What is good QoS/QoE for Space Video

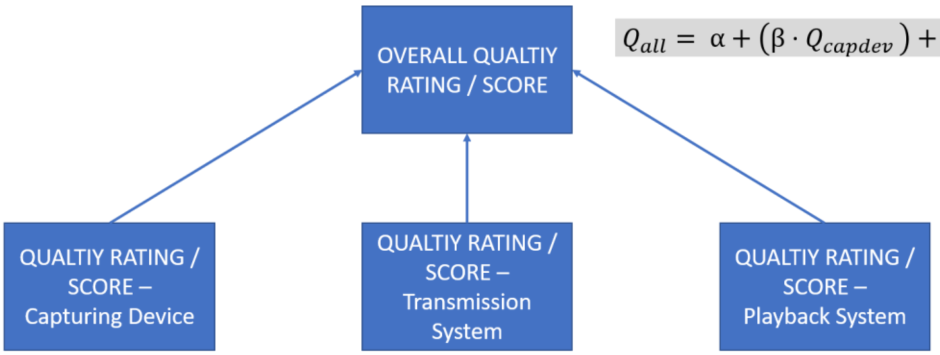
Another essential aspec is to determine what good or excellent video quality for space video is. Again, this is depended on the scenario and relevant stakeholders, so it needs to be determined for each separately.
Furthermore, one should be clear about the goal; should QoS, QoE, or both be maximized, or should a threshold for acceptable quality be found.

So, the first practical task would be to generate a database for each scenario, with a broad range of content and technical details. These databases are then rated by individuals regarding the video quality. The rating together with the technical details of the source material, can then be used as a benchmark for later, quality model usage (commercial, self-developed etc.).

The second task is to investigate which recommendations, models, and metrics can be applied or adjusted to enable statements about the video quality.
The third task is to validate the chosen approach and make recommendation for the use cases and scenarios.

6. A possible approach to model the Video Quality for Space Video

An approach could be to divide "Video Quality" into independent "Video Quality Parts", one for each subsystem of the overall video transmission system from the capturing device to the playout on a screen. If necessary these subsystems may be subdivided into smaller components. These videos parts can be weighted and added to the overall/integral video quality of the system as depicted in the figure below:



$$Q_{all} = \alpha + (\beta \cdot Q_{capdev}) + (\gamma \cdot Q_{trans}) + (\delta \cdot Q_{play})$$

$$Q_{capdev} = \alpha + (\beta \cdot Imp_{Lense}) + (\gamma \cdot Imp_{Sensor}) + (\delta \cdot Imp_{Format}) + (\ldots)$$

$$Q_{trans} = \alpha + (\beta \cdot Imp_{bitrate}) + (\gamma \cdot Imp_{packet\ loss\ rate}) + (\delta \cdot Imp_{transcode}) + (\ldots)$$

$$Q_{play} = \alpha + (\beta \cdot Imp_{transcode}) + (\gamma \cdot Imp_{display}) + (\delta \cdot Imp_{Format}) + (\ldots)$$

Here, the quality scores are calculated separately. The "Imp" are impairment factors. Every step in the processing chain has the potential to impair the video signal. All the impairment factors are added and weighted, depending on the severeness and/or the importance of the impairment.

The calculation of these impairment factors can be very complex. To elaborate on that, an example from the ITU-T Recommendation G.1070 [8] on video quality is presented in the following:

Basic video quality affected by coding distortion, I_{coding}

The basic video quality affected by coding distortion I_{coding} is expressed as:

$$I_{coding} = I_{Ofr} \exp\{\frac{\left(\ln(Fr_v) - \ln(O_{fr})\right)^2}{2D_{Frv}^2}\}$$

The O_{fr} is an optimal frame rate that maximizes the video quality at each video bit rate (B_{rV}) and is expressed as:

$$O_{fr} = v_1 + v_2 B_{rv}, \qquad 1 \le O_{fr} \le 30, v_1 and\ v_2: const$$

where $F_{rV} = O_{fr}$, $I_{coding} = I_{Ofr}$, I_{Ofr} represents the maximum video quality at each video bit rate (B_{rV}) and is expressed as:

$$I_{Ofr} = v_3 - \frac{v_3}{1 + \left(\frac{B_{rv}}{v_4}\right)^{v_5}} \ ,0 \le I_{Ofr} \le 4 \ , v_3, v_4\ and\ v_5: const$$

D_{FrV} represents the degree of video quality robustness due to frame rate (F_{rV}) and is expressed as:

$$D_{FrV} = v_6 + v_7\, B_{rV} \ ,0 < D_{FrV} \ , v_6\ and\ v_7: const$$

Coefficients v1, v2, ..., and v7 are dependent on the codec type (CT), video format (VF), key frame interval (KFI) and video display size (VDS).

Finally, here is a list of Recommendations, Reports and Standard from the ITU-T and ESTI. These are regarded as starting point to investigate their applicability (full or partial) for space video quality assessment.

Recommendations:

1. ITU-T Recommendation P.800.1 (2016), Mean opinion score (MOS) terminology.
2. ITU-T Recommendation P.911 (1998), Subjective audiovisual quality assessment methods for multimedia applications.
3. ITU-T Recommendation P.913, Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment
4. ITU-T Recommendation P.918, Dimension-based subjective quality evaluation for video content
5. ITU-T Recommendation P.931, Multimedia communications delay, synchronization and frame rate measurement
6. ITU-T Recommendation P.1201.1 (2012), Parametric non-intrusive assessment of audiovisual media streaming quality – Lower resolution application area.
7. ITU-T Recommendation P.1201.2 (2012), Parametric non-intrusive assessment of audiovisual media streaming quality – Higher resolution application area.
8. ITU-T Recommendation P.1202 (2012), Parametric non-intrusive bitstream assessment of video media streaming quality.
9. ITU-T Recommendation P.1202.1 (2012), Parametric non-intrusive bitstream assessment of video media streaming quality – Lower resolution application area.

10. ITU-T Recommendation P.1203 (2017), Parametric bitstream-based quality
11. ITU-T Recommendation P.1204 (2020), Video quality assessment of streaming services over reliable transport for resolutions up to 4K.
12. ITU-T Recommendation P.1204.1, Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to transport information (in development)
13. ITU-T Recommendation P.1204.2, Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to video frame information (in development)
14. ITU-T Recommendation P.1204.3 (2020), Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to full bitstream information.
15. ITU-T Recommendation P.1204.4 (2020), Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to full and reduced reference pixel information.
16. ITU-T Recommendation P.1204.5 (2020), Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to transport and received pixel information. assessment of progressive download and adaptive audiovisual streaming services over reliable transport.
17. ITU-T Recommendation G.1050, Network model for evaluating multimedia transmission performance over Internet Protocol
18. ITU-T Recommendation G.1070, Opinion model for video-telephony applications
19. ITU-T Recommendation G.1071, Opinion model for network planning of video and audio streaming applications
20. ETSI TR 102 805-3 V1.1.1, User Group; End-to-end QoS management at the Network Interfaces; Part 3: QoS informational structure
21. ETSI TS 103 294 v1.1.1, Speech and multimedia Transmission Quality (STQ); Quality of Experience; A Monitoring Architecture
22. ETSI TR 103 559 V1.1.1, Speech and multimedia Transmission Quality (STQ); Best practices for robust network QoS benchmark testing and scoring
23. ETSI TS 102 464 V1.2.1, Satellite Earth Stations and Systems (SES); Broadband Satellite Multimedia (BSM); Interworking with DiffServ QoS
24. ETSI EG 202 934 V1.1.2, User Group; The assessment of the overall Quality of Services (QoS) as perceived by the users; Definition of QoS indexes for all the customer relationship stages

Bibliography

[1] CCSDS Recommended Standard 766.2-B-1, ”Blue Book – Voice and Audio Communications”, Consultative Committee for Space Data Systems, USA-Washington, D.C., 11/2017

[2] U. Jekosch, ”Voice and Speech Quality Perception: Assessment and Evaluation”. Springer, GER-Berlin, 2005.

[3] S. Möller and A. Raake, Eds., ”Quality of Experience: Advanced Concepts, Applications and Methods”, 1st ed. Springer, GER-Heidelberg, 2014.

[4] ITU-T Rec. E.800, ”Definitions of Terms related to Quality of Service”, International Telecommunication Union, CH-Geneva, 09/2008.

[5] ITU-T Rec. P.10/G.100, ”Vocabulary for Performance, Quality of Service and Quality of Experience”, International Telecommunication Union, CH-Geneva, 11/2017.

[6] ETSI TS 103 294, ”Speech and Multimedia Transmission Quality (STQ); Quality of Experience; A Monitoring Architecture”, European Telecommunications Standards Institute, FRA-Sophia Antipolis, 12/2014.

[7] F. Schiffner, ”Dimension-based Quality Analysis and Prediction for Videotelephony”, T-Labs Series in Telecommunication Services, Springer, CH-Cham, 2020.

[8] ITU-T Rec. G.1070, ”Opinion model for video-telephony applications”, International Telecommunication Union, CH-Geneva, 06/2018.