ETSI TR 102 493 V1.3.1 (2017-07)

TECHNICAL REPORT

Speech and multimedia Transmission Quality (STQ); Guidelines for the use of Video Quality Algorithms for Mobile Applications

Reference

RTR/STQ-00211m

Keywords

QoS, telephony, video

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 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is 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 IPR Policy, no investigation, 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.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ).

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.

1 Scope

The present document gives guidelines for the use of video quality algorithms for the different services and scenarios applied in the mobile environment.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ETSI TS 126 233 (V13.0.0): "Universal Mobile Telecommunications System (UMTS); LTE; Transparent end-to-end Packet-switched Streaming service (PSS); General description (3GPP TS 26.233 version 13.0.0 Release 13)".

[i.2] ETSI TS 126 114: "Universal Mobile Telecommunications System (UMTS); LTE; IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction (3GPP TS 26.114)".

[i.3] Void.

[i.4] Recommendation ITU-T P.1201.1: "Parametric non-intrusive assessment of audiovisual media streaming quality - Lower resolution application area".

[i.5] Recommendation ITU-T P.1201.2: "Parametric non-intrusive assessment of audiovisual media streaming quality - Higher resolution application area".

[i.6] Recommendation ITU-T P.1202.1: "Parametric non-intrusive bitstream assessment of video media streaming quality - Lower resolution application area".

[i.7] Recommendation ITU-T P.1202.2: "Parametric non-intrusive bitstream assessment of video media streaming quality - Higher resolution application area".

[i.8] Recommendation ITU-T P.1203.1: "Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport - video quality estimation module".

[i.9] Recommendation ITU-T P.1203.2: "Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport - audio quality estimation module".

[i.10] Recommendation ITU-T P.1203.3: "Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport - Quality integration module".

[i.11] Recommendation ITU-T J.343.1: "Hybrid-NRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of encrypted bitstream data".

[i.12] Recommendation ITU-T J.343.2: "Hybrid-NR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of non-encrypted bitstream data".

[i.13] Recommendation ITU-T J.343.3: "Hybrid-RRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a reduced reference signal and encrypted bitstream data".

[i.14] Recommendation ITU-T J.343.4: "Hybrid-RR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a reduced reference signal and non-encrypted bitstream data".

[i.15] Recommendation ITU-T J.343.5: "Hybrid-FRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a full reference signal and encrypted bitstream data".

[i.16] Recommendation ITU-T J.343.6: "Hybrid-FR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a full reference signal and non-encrypted bitstream data".

[i.17] Recommendation ITU-T J.246 (2008): "Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference".

[i.18] Recommendation ITU-T J.247 (2008): "Objective perceptual multimedia video quality measurement in the presence of a full reference".

[i.19] Recommendation ITU-T J.341 (2016): "Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

bitstream model: computational model that predicts the subjectively perceived quality of video, audio or multimedia, based on analysis of the payload and transport headers

hybrid model: computational model that predicts the subjectively perceived quality of video, audio, or multimedia, based on the media signal and the payload and transport headers

live Streaming: streaming of live content e.g. web cam, TV programs, etc.

parametric model: computational algorithm that predicts the subjectively perceived quality of video, based on transport layer and client parameters

perceptual model: computational algorithm that aims to predict the subjectively perceived quality of video, based on the media signal

streaming on demand: streaming of stored content e.g. movies

3.2 Abbreviations

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

AVC	Advanced Video Coding
BLER	BLock Error Rates
CPU	Central Processing Unit
DCT	Discrete Cosine Transform
FR	Full Reference Algorithm

HD	High Definition

NOTE: 1 280 x 720 pixels, fullHD 1 920 x 1 080 pixels.

HEVC	High Efficiency Video Coding
IMS	IP Multimedia Subsystem
IP	Internet Protocol
ITU	International Telecom standardization Union
JPEG	Joint Photographic Expert Group (Standard)
MM	Multimedia
MOS	Mean Opinion Score
MPEG	Moving Picture Expert Group (Standard)
MPEG-TS	MPEG Transport Stream
MTSI	Multimedia Service for IMS
NR	No Reference Algorithm
PLR	Packet Loss Rates
PSNR	Peak Signal Noise Ratio
RR	Reduced Reference
RTP	Real Time Protocol
SD	Standard Definition
TCP	Transport Control Protocol
TV	Television
UDP	User Datagram Procotol
VGA	Video Graphics Adapter
VHS	Video Home System

4 General

Video quality assessment has become a central issue with the increasing use of digital video compression systems and their delivery over mobile networks. Due to the nature of the coding standards and delivery networks the provided quality will differ in time and space. Thus, methods for video quality assessment represent important tools to compare the performance of end-to-end applications.

The present document sets the guidelines of video quality algorithms applicable for mobile applications and the scenarios of their application. Any eligible algorithm needs to predict the quality perceived by the user using mobile terminal equipment. The goal is to have one or more objective video quality measurement algorithms, which predict the video quality as perceived by a human viewer, which is in conformance with the minimum requirements list given in the present document.

ITU-T has approved many different algorithms for objective prediction of visual quality in the last years. They can be differentiated in algorithms based on image analysis (the actual image is input to the algorithm) and bitstream-based measures (the IP stream is input of the model), and so-called hybrid models which combine image analysis with meta-information from the bitstream. These models have different scopes and limitations, they require different input information and result in different predictions accuracy.

For image based analysis ITU recommends a reduced reference algorithm in Recommendation ITU-T J.246 [i.17] up to VGA resolution, multiple full-reference algorithms as in Recommendation ITU-T J.247 [i.18] for all video resolutions up to VGA and the algorithm in Recommendation ITU-T J.341 [i.19] for HD resolutions.

Bitstream based models are described in Recommendations ITU-T P.1201 and P.1202 series ([i.4] to [i.7]); there are differentiations for individual resolutions and for encrypted and non-encrypted bitstreams.

In 2015 ITU approved a set of hybrid models in the Recommendation ITU-T J.343 series ([i.11] to [i.16]), where J.343.1 [i.11] and J.343.2 [i.12] are no-reference hybrid models, J.343.3 [i.13] and J.343.4 [i.14] are reduced reference hybrid models and J.343.5 [i.15] and J.343.6 [i.16] are full-reference hybrid models. The odd suffix stands for models to be applicable for encrypted and non-encrypted bitstreams, while the models having an even suffix are only applicable to non-encrypted bitstreams.

It is common to all services treated in the present document that quality as seen from the user's perspective depends on the server and client applications used. For example, is has to be expected that under the same network conditions, two different video streaming clients will exhibit different video quality due to differences in the way these clients use available bandwidth. Therefore, for full validation of tools type and version of clients used has to be fully documented and are seen as part of the information needed to reproduce and calibrate measurements.

NOTE: The present document focuses on those visual continuous media reproductions where the source and the player are connected via a (mobile) telecommunication network rather than the replay of a clip that has been completely stored on the same device as the player and is replayed from there.

5 Services

5.0 Introduction

The aspect of video quality is of interest wherever there are services where the transfer of moving pictures or still images is involved. Three major fields of transferring video content can be identified that make use of packet switched and circuit switched services.

Table 1: Requirement profiles of the services

Application	Symmetry	Data rates	One Way Delay	Lip-sync
Video telephony	Two-way	32 kbps to 2 Mbps	< 150 ms preferred < 400 ms limit	< 80 ms
Streaming	One-way	32 kbps to 10 Mbps	< 10 s	
Conversational Multimedia	Two-way		< 150 ms	Mutual service dependency, echo

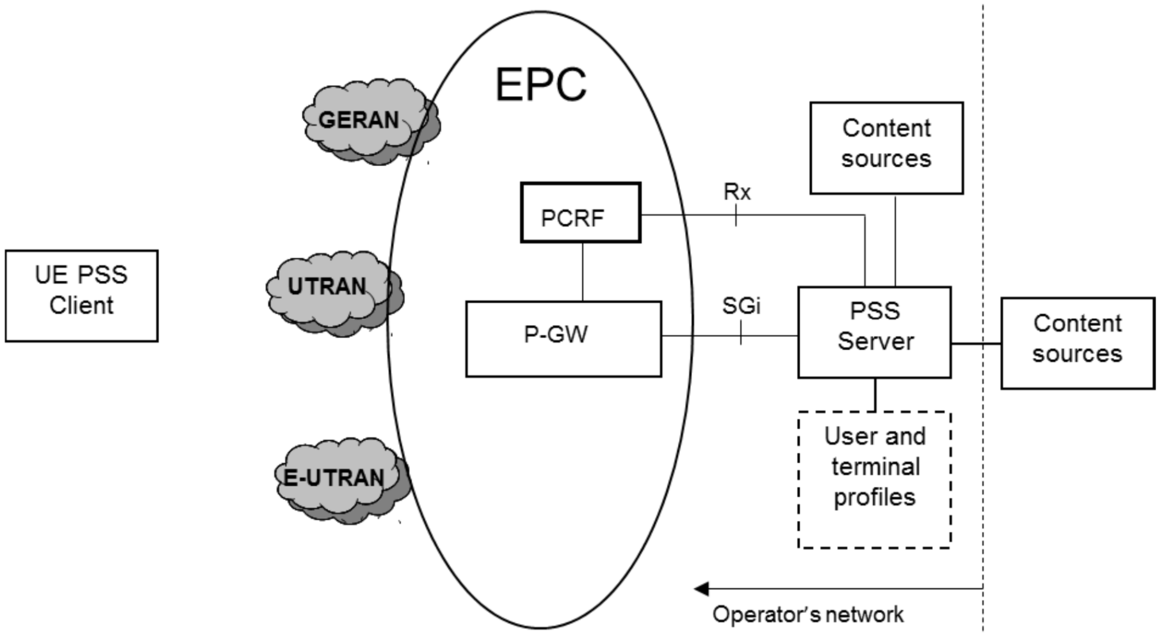

Figure 1: Streaming (ETSI TS 126 233 [i.1])

5.1 Streaming

Streaming refers to replay of media streams like audio and video in a continuous way while those streams are being received by the client over a data network. The client plays the incoming multimedia stream from a buffer in which the packets are stored after arrival.

Streaming accounts for a large percentage of the data network traffic. Typical applications can be classified into on-demand and live information delivery applications. Examples of the first group are video-on-demand applications like YouTube™. Live delivery of radio and television programs is an example of the second category.

NOTE: YouTube™ is the trade name of a product supplied by Google. This information is given for the convenience of users of the present document and does not constitute an endorsement by ETSI of the product named.

5.2 Conversational Multimedia

Multimedia services combine two or more media components within a call. The service where two or more parties exchange video, audio and text and maybe even share documents is a multimedia service. This is a peer-to-peer set up in which one party acts as the source (server) and the other as client(s) and vice versa in real time. Another example of a new multimedia conversational service is the 3GPP standardized MTSI service [i.2].

5.3 Video Telephony

Video telephony is a full-duplex system, carrying both video and audio and intended for use in a conversational environment. In principle the same delay requirements as for conversational voice will apply, i.e. no echo and minimal effect on conversational dynamics, with the added requirement that the audio and video have to be synchronized within certain limits to provide "lip-synch".

6 QoS Scenarios

6.0 General

The different services that are making use of video can be delivered in a variety of ways and situations. To obtain the full picture of the quality of these services they need to be tested accordingly. However for practical purposes and general feasibility, key scenarios need to be identified to facilitate video quality measurements.

6.1 Measurement Scenarios

The key scenarios are live streaming, streaming on demand, video telephony and conversational multimedia. These services can be tested by drive test or in a static fashion.

The algorithms for estimating video and audiovisual quality can be classified depending on:

- Type of input:
    - Perceptual (access to the video signal).
    - Hybrid (access to both the video signal and either the transport layer payload or the transport header information).
    - Bitstream (access to the transport layer payload, but not the video signal).
    - Parametric (access to transport header, client information, and knowledge about used codecs).

NOTE: The accessibility of the needed information depends on presence of encryption and its depth.

- Access to reference video: The algorithm models that are used are:
    - Full reference model (FR).
    - No reference model (NR).

- Media types: An algorithm can estimate:
  - Video quality only.
  - Audiovisual quality (taking into account the combined effect of audio and video quality).

Table 2: Key scenarios and model applicability for video quality algorithm assessment

	Live streaming	Streaming on Demand	Video Telephony	Conversational MM
FR perceptual	Require pre-stored source - normally not applicable for live streaming.	Applicable. Require pre-stored source.	Applicable. Require pre-stored source.	Applicable. Require pre-stored source.
NR perceptual	Applicable. Might have bad performance when video contains artefact-like content.	Applicable. Might have bad performance when video contains artefact-like content.	Applicable. Might have bad performance when video contains artefact-like content.	Applicable. Might have bad performance when video contains artefact-like content.
FR hybrid	Require pre-stored source - normally not applicable for live streaming.	Applicable. Require pre-stored source.	Applicable. Require pre-stored source.	Applicable. Require pre-stored source.
NR Hybrid	Applicable.	Applicable.	Applicable.	Applicable.
Bitstream	Applicable.	Applicable.	Applicable.	Applicable.
Parametric	Applicable.	Applicable.	Applicable.	Applicable.

6.2 Other scenarios

There is a further approach of video testing that does not focus on the perceptual quality of a delivered video but on the pure availability (delivery) of the desired content in real time. This is referred to as live verification or live monitoring. Like in the previous clause all four scenarios can be tested with all models. However due to the nature of the NR, parametric and bitstream models they are more suitable for that purpose.

7 Requirements for test systems for mobile networks

7.0 General

Testing of mobile networks is a special field of application for a video quality algorithm. To be actually applicable for e.g. drive testing any algorithm should fulfil the following requirements.

7.1 Sequence and observation length

Since one aspect of mobile network testing is to georeference the results to identify areas with less than optimal quality, the algorithm should be capable to provide data for a reasonable resolution. Therefore it should be capable of assessing sequences of a period of 8 seconds to 30 seconds (comparable with listening quality).

The length of a Video Telephony call and video streaming can vary between a couple of seconds and several hours. For video streaming sessions where the quality is degraded by rebuffering, the sequence length should be in the range 15 seconds to 30 seconds to be able to estimate the quality for such degradations.

Estimating quality for sequences longer than 30 seconds may be done by collecting and aggregating the results of a sequence of short samples. The way of aggregation needs to be determined.

7.2 Content

The algorithm should be capable of assessing the quality of all visual content that is (can be) delivered over mobile networks. E.g.:

1) Video conferencing.

2) Movies, movie trailers.

3) Pictures/Still images.

Regarding 3) it is required that the algorithm can process pictures of the type of content delivered as moving picture (1,2) and in addition still images and maps. When using a perceptual or hybrid algorithm, the test set-up should include a variety of content and the final quality should be the average of all used contents. A parametric quality model normally directly estimates the average quality for typical video or audiovisual content.

7.3 Algorithms

7.3.1 Image-based algorithms

7.3.1.0 General

Image based algorithms require the access to the decoded video images for analysis. To receive a quality prediction for a video as seen by a subscriber, preferably the actually displayed images at the end-user device are analysed by the algorithms.

In general image based algorithms can be sub-classified in 'full-reference', 'reduced-reference' and 'no-reference' approaches. Full-reference models require the access to a full reference video; the quality estimation is usually based on a pixel- and frame-wise comparison of the captured video to the reference. Reduced-reference models do require some meta information of the reference (input) video, but not the full bitmaps. No-reference models do not require any information about the reference. The evaluation of the video is made by the image analysis without comparison to a known reference signal.

In this class fall also so-called hybrid models, where the image analysis is extended by meta information taken from the bitstream. These models require the incoming IP bitstream as second input. In case the required information taken from the bitstream is very restricted, the main analysis is still based mainly on image evaluation.

Because image based algorithms in general are assessing the decoded images, they are transparent to encryption and widely applicable to used protocols and containers. The encryption and depackaging is handled by the actual video player and decoded at the device. The algorithm accesses the displayed video after decoding as a user does.

Compared to bitstream models, the image analysis considers artefacts from the entire transmission chain independent of potential transcodings and/or re-packagings in the transport domain e.g. IP layer. For full-reference methods the quality of the video source is under control. High quality source videos are mandatory, otherwise the quality prediction will most likely fail. No-reference models do not require a dedicated reference video and take into account the source quality too.

The restriction for full- and reduced reference approaches is the knowledge of the reference video. Usually, a known reference video is sent into the video channel or uploaded to a video server platform and streamed down to the users' device and analysed by the algorithm. Therefore full- and reduced reference methods cannot be used for live video evaluation.

Image based models are typical end-point measures. They require access to the decoded video that is typically at a user's device as e.g. a smart phone, or to be connected to it, or at the output of a hardware video player, for example a set top box.

Image based models reflect the user's perception at best, because they evaluate typically the image the user really sees. They conduct active testing and acquire a test connection. Bitstream models instead apply a general video player model; they do not access the actual picture and are therefore more usable for general analysis and passive video network monitoring.

7.3.1.1 Image-based algorithms with access to a reference

Full-reference methods are comparing a received and potentially degraded video signal with a high-quality uncompressed video source. This comparison is usually done per frame and quasi pixel-wise. Potential differences of the video to the reference are considered as degradations and weighted according to their perceptual impact for a viewer.

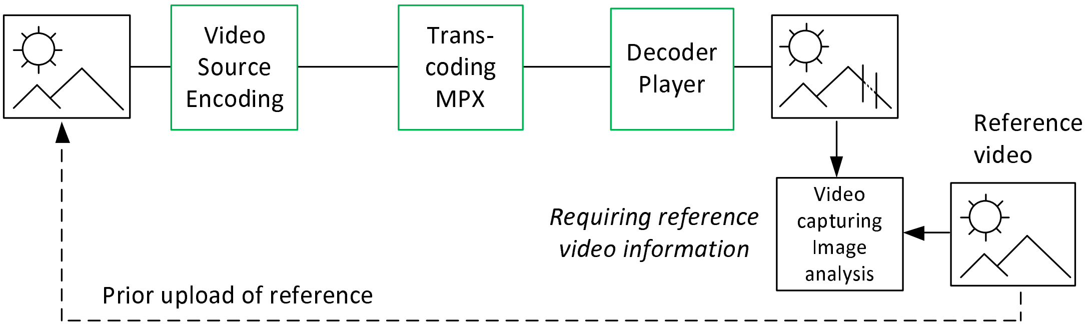
Video Source Encoding
Trans-coding MPX
Decoder Player
Reference video
Requiring reference video information
Video capturing Image analysis
Prior upload of reference

Figure 2: Scheme of an image based full-reference video measurement approach

Full reference methods are considered as very accurate because of the detailed comparison to a high-quality reference signal. They are well applicable for evaluation of coding and transmission distortions, where the video can be digitally inserted at the sending side and digitally captured at the receiving side, preferably at the same image resolution, without possible re-scaling to a different display resolution.

Full-reference methods require access to the high-quality reference for the comparison. It can be achieved by prior uploading of a reference video and streaming it on demand in a dedicated test to the probe, where the algorithm is applied. An alternative possibility is the capture of the source signal while sending and transferring the video by other means to the probe, where the algorithm is running. This requirement of accessing the reference makes full-reference algorithms unusable for live video as e.g. TV broadcasting or live captured video transmissions.

There are also image based reduced-reference algorithms, where instead of the full images of the reference video, the algorithm only requires some specified characteristics of the reference video. The restrictions remain the same regarding access to the reference for computing the characteristics and therefore its restriction for live video.

For both types of algorithms there are special implementations available, which are taking meta information from the IP-stream into account for an enhanced quality prediction. Those image based models are called 'full reference hybrid models' and 'reduced reference hybrid models', respectively.

7.3.1.2 Image based no reference perceptual algorithms

No reference perceptual algorithms evaluate the quality based on the video signal played back at the receiver, without access to a dedicated high quality reference signal.

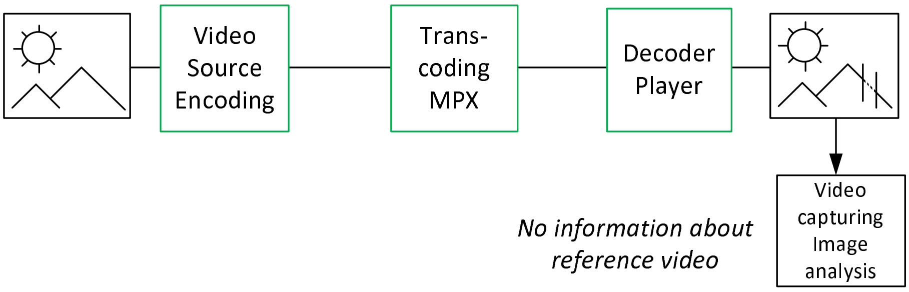
Video Source Encoding
Trans-coding MPX
Decoder Player
No information about reference video
Video capturing Image analysis

Figure 3: Scheme of an image based no-reference video measurement approach

No-reference algorithms are transparent to re-scaling; they only assess the received video in its delivered or displayed resolution. Like all image based algorithms, image based no reference methods are also transparent to encryption, transcoding and re-packaging.

While for full- and reduced-reference methods the user can choose the reference video and its content, for image-based no reference models erroneous evaluation caused by particular content is to be avoided. For example , that artefact-like content is not to be confused with real artefacts, e.g. graphical squares or artificial blurring. Furthermore that black scenes received or still images are not considered as freezing and do not produce high MOS scores if the source of the videos was not black or a still image respectively. For valid and reproducible predictions the analysis of natural content with a common amount of motion and spatial complexity is recommended.

Also for image based no-reference methods, special implementations make use of IP meta information to enhance the prediction accuracy. These image based models are called 'no reference hybrid models'.

No reference algorithms are independent from a dedicated video server providing reference video samples and in general have the capability to score live video.

7.3.2 Bitstream algorithms

A bitstream algorithm uses the IP bitstream to estimate the perceived quality. It does not use the decoded video signal at the receiver or the reference video sequence. The quality-relevant information is derived from the information in the IP payload (if accessible) and from IP meta information. Based on the presence of encryption, the analysis remains restricted to the available information and no payload analysis can be conducted. For encrypted bitstreams or encrypted payload, only meta information can be used, for unencrypted content, also the video payload information can be accessed and analysed.

Bitstream models only analyse the current bit stream, they cannot consider distortions caused by prior compressions and happened in pre-switched channel. They do not have access to information before potential transcoding and/or re-packaging. Bitstream based models have also only very limited means to consider low source quality. The algorithm does not use the received video as input, but can still give a score depending on the content. Analysis of the bitstream can indicate what type of content the signal consists of in case the payload is accessible.

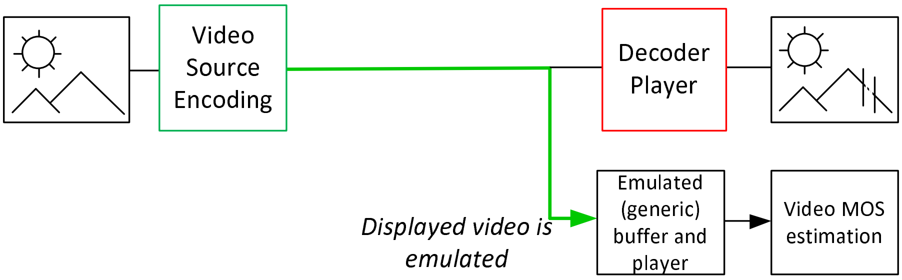
Video Source Encoding
Decoder Player
Displayed video is emulated
Emulated (generic) buffer and player
Video MOS estimation

Figure 4: Scheme of a bitstream based video measurement approach

Because bitstream models parse all or part of the bitstream, they are restricted to defined protocols, container types and codecs as advised in the scope of the individual specification.

Bitstream models do not analyse the actual video as visible on the screen, they are simulating the video buffer, the player and the decoder characteristics on their own, algorithm-internal, general model of these components. As they are not using the actually displayed video to the viewer, those models can be applied for passive, non-intrusive network monitoring at mid-point measurement points.

7.3.3 Parametric algorithms

A parametric media quality algorithm estimates the perceived quality based on measurement parameters, but not based on the actual video and audio signals. Typical input to a parametric algorithm is information about codec, coded bitrate, transport errors and client information about buffering.

A parametric algorithm is trained to estimate the quality for typical and average video content, and some algorithms will give the same score for a given codec, bit rate and transport error situation independent of the video content. Some algorithms might be able to take some content aspects into account.

A parametric algorithm is able to score live video, since detailed information about the source video is not required. The algorithm typically requires information about codec and coded bit rate. This type of algorithm may still be applicable when only an encrypted bitstream is available. Parametric models can be considered bitstream models with very restricted information on bitstream.

7.3.4 Supported Video Codecs

Video codecs may include but are not limited to:

- H.263.
- MPEG4 - part 2.
- H.264/AVC.
- HEVC (H.265).
- VP9.

7.3.5 Calculation time

The calculation time should be as short as possible without any negative impact on the accuracy of the results. The calculation time should be shorter than the actual sequence duration.

7.4 Container schemes

Container schemes that will be used may include, but are not limited to:

- MPEG4.
- 3GPP.
- WebM.
- Proprietary schemes.

7.5 Output

Given the complexity of videos and the degrees of freedom of errors each assessment can have a complex result. However there should be one overall value for each assessment that allows an easy comparison of results gathered under different conditions. Therefore the algorithms output should be one value on the MOS scale from 1 to 5 with a resolution of two decimal digits for each rated video sequence. The score 1 represents *bad* and the score 5 represents *excellent* quality. Note that all algorithms are tuned against a number of subjective tests. Since subjective tests are done with humans and probably with slightly different test set-ups, the scores from two subjective tests will not be exactly the same. Hence, two models trained on different subjective test databases will not give exactly the same score.

8 Standardized algorithms for video quality prediction

8.0 General

There are multiple standardized algorithms for video quality prediction in force. The most recent ones are also capable of HD resolutions.

8.1 Bitstream based and parametric algorithms

Currently standardized bitstream based and parametric models are:

- Recommendations ITU-T P.1201.1/.2 Parametric, non-intrusive audiovisual media streaming quality [i.4] and [i.5].
- Recommendations ITU-T P.1202.1/.2 Parametric non-intrusive bitstream assessment of video media streaming quality [i.6] and [i.7].
- Recommendations ITU-T P.1203.1/.2/.3 Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport [i.8] and [i.9].

The standardized bitstream and parametric models predict visual quality, audio quality and audio-visual quality.

There are several sub-standards in these Recommendations describing individual models for lower (SD) (Recommendations ITU-T P.1201.1 [i.4] and P.1202.1 [i.6]) and for higher (HD) resolution (Recommendation ITU-T P.1201.2 [i.5] and P.1202.2 [i.7]) and within the Recommendations flavours for encrypted and non-encrypted videos.

Recommendation ITU-T P.1203 is separated in sub-standards for video, audio and the audiovisual integration module [i.8], [i.9] and [i.10]. Recommendation ITU-T P.1203.1 [i.8] is restricted to reliable content transmission as in TCP protocols, while Recommendations ITU-T P.1201 ([i.4] to [i.5]) and P.1202 ([i.6] to [i.7]) are verified and recommended for unreliable content transmission (transmission over RTP/UDP for lower resolution, and transmission over MPEG2-TS/RTP/UDP or MPEG2-TS/UDP for higher resolution).

The mentioned bitstream models are all 'no-reference' and even non-intrusive models, they do not require a known test video. They can be applied to any video including live video as typical for no-reference measures.

8.2 Image based algorithms

Currently standardized image based and HD capable models are ITU J.341 Image based full-reference algorithm for HD resolution and the hybrid models:

- Recommendations ITU-T J.343.1/.2 Image based no-reference hybrid algorithm for up to HD resolution [i.11] and [i.12].
- Recommendations ITU-T J.343.3/.4 Image based reduced-reference hybrid algorithm for up to HD resolution [i.13] and [i.14].
- Recommendations ITU-T J.343.5/.6 Image based full-reference hybrid algorithm for up to HD resolution [i.15] and [i.16].

The odd/even numbered sub-standards differentiate in models applicable for encrypted (odd) and non-encrypted (even) bitstreams. The hybrid models can be used for unreliable content transmission (transmission over RTP/UDP and MPEG-TS/RTP/UDP) and for reliable content transmission if the necessary model input data is accessible.

Annex A: Algorithms

A.0 Introduction

Existing QoS indicators such as peak signal-to-noise ratio (PSNR) or network statistics like Packet Loss (PLR) and BLock Error Rates (BLER) are not sufficient to measure the quality that a typical subscriber perceives. The reasons for this are two-fold:

1) The bits in a multimedia bit stream have different perceptual importance. Depending on which part of the bit stream is affected by errors or losses, the same amount of data losses can have significantly different perceptual effects on the presented multimedia content.

2) The human visual and auditory systems process information in an adaptive and non-uniform fashion. This means that the annoyance of artefacts depends on the type of artefact as well as the characteristics of the content in which they occur.

These facts call for quality metrics, which assess multimedia content in a similar fashion as is done by the human visual and auditory systems.

The new objective measurement methods analyse the video signal in the video image space employing knowledge of the human visual system. These methods apply to algorithms that measure image quality usually based on the comparison of the source and the processed sequences. The challenge of developing techniques for the quality estimation of video compression systems partly lies in the fact that compression algorithms and delivery over mobile networks introduce new video impairments, impairments that strongly depend on the levels of detail and motion in the scenes. Therefore traditional assessment methods, which use static test signals, are inadequate to measure the performance of modern video compression systems.

Nevertheless the video algorithms working with these new methods need to be validated for real applications. The basis for this validation will be the MOS obtained from controlled subjective tests for a set of test sequences given by human watcher. Depending on the type of validation the results of the objective and the subjective tests will be confronted. The performance of objective models will be based on the accuracy of the prediction of the MOS. The goal for any video quality algorithm is to predict the subjective rating as good as possible.

A.1 Measurement Methodologies

A.1.0 Introduction

When designing algorithms or *metrics* to assess perceptual quality, three basic methodologies can be chosen (most arguments hold equally for Audio). Each methodology has its advantages and limitations. The objectives underpinning the measurements should help decide which methodology is most suitable for a given measurement scenario.

Traditional methods are able to accurately measure and assess analogue impairments to the video signal. However, with the introduction and development of digital technologies, visually noticeable artefacts appear in ways that are different from analogue artefacts. This change has led to the need for new objective test methods.

A.1.1 Full Reference Approach (FR)

The FR technique is based on a comparison of the original content (*Reference*) with what is received at the terminal (*Processed*).

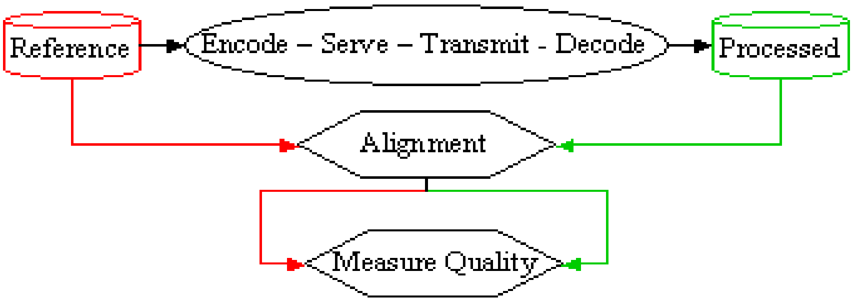

Figure A.1: Full Reference methodology

FR metrics compute the difference between a Reference and its corresponding Processed video. This difference is then analysed in view of characteristic signatures such as blur or noise. A classic FR metric used widely in the literature is PSNR (Peak Signal to Noise Ratio). Perceptual FR metrics can be made extremely sensitive to subtle degradations and can be designed to detect very specific artefacts.

In order to use the FR approach, the Reference has to be available for the processing.

In FR methods it is often necessary to separately register the reference and processed sequences. Registration is a process by which the reference and processed video sequences are aligned in both the temporal and spatial domains. The degree to which alignment is necessary can differ depending on the functionality of a particular model, and it is possible that FR models may include alignment as an integral part of the measurement method or even not require registration at all.

Where registration is required, the alignment algorithm will need to have access to both the reference and processed content. This has two important implications:

a) Resources to store the Processed content have to be made available.

b) Analysis results are not immediately available (see table A.1, line "Real time").

In this sense, FR techniques are invasive and are limited to relatively short sequences. Please note that no compression should be used during capture and storage of the Processed sequence.

A.1.2 No Reference Approach (NR)

The NR technique is based on an analysis of the Processed content without any knowledge of the Reference.

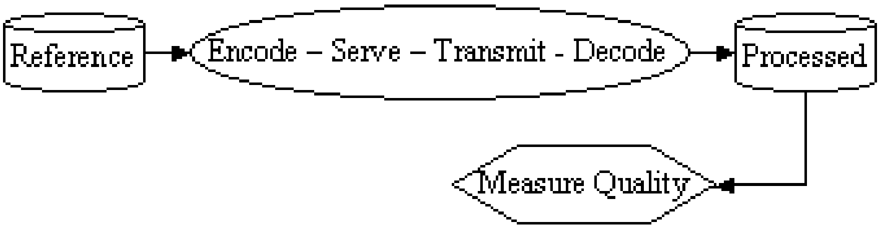

Figure A.2: No Reference methodology

NR metrics depend on a preset scale. This scale should be defined by the quality range that can be expected. This, for video, is principally determined by the following factors:

- Encoder target bit rate.
- Codec type.
- Frame size.
- Frame rate.

NR metrics measure characteristic impairments through feature extraction and pattern matching techniques. The types and characteristics of the target features are chosen to have a high perceptual impact and need to be carefully tuned and weighted according to the characteristics of the human visual system.

NR metrics provide a general indication as to the level of target impairments. Under certain circumstances, they can be misled by content containing characteristics which look like an impairment.

EXAMPLE: An image of a chessboard may trigger a metric targeting blockiness to measure a high degree of impairment. If a video sequence contains still images, a metric targeting jerkiness may indicate bad quality.

NR metrics do not require alignment nor do they depend on the entire Processed to be available at the time of analysis. Thus they are ideally suited for in-service quality measurement of live video streaming or video telephony. They enable live-service monitoring measurement solutions for any video at any point in the content production and delivery chain. NR metrics are particularly useful for monitoring quality variations due to network problems, as well as for applications where SLAs need to be enforced.

A.1.3 Reduced Reference Approach (RR)

The RR technique tries to improve on FR by reducing computational and resource requirements at the point of analysis.

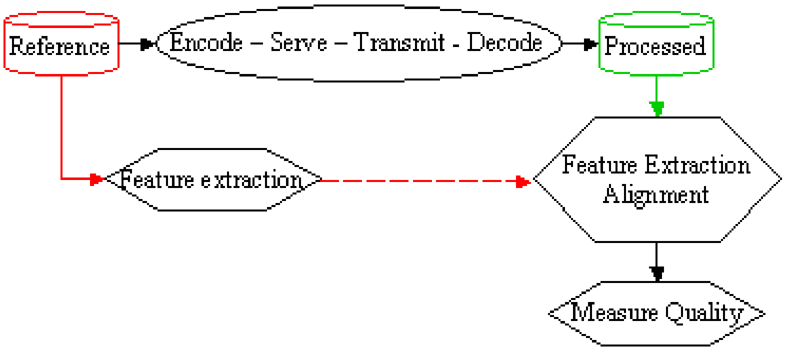

Figure A.3: Reduced Reference methodology

The reduced-reference approach lies between the extremes of FR and NR metrics. RR metrics extract a number of representative features from the reference video (e.g. the amount of motion or spatial detail), and the comparison with the Processed video is then based only on those features. This makes it possible to avoid some of the pitfalls of pure no-reference metrics while keeping the amount of reference information manageable. Nonetheless, the issues of reference availability and alignment remain.

To take the full advantage of the RR approach the information extracted from the reference needs to be transmitted together with the test clip. In doing that the information is taking away bandwidth of the channel that is to be measured. Therefore the RR model appears not to be suitable for mobile video quality measurements.

A.1.4 Comparison of FR and NR Approaches

Focussing on the full reference and the no reference perceptual model the two approaches can be compared in various aspects.

Table A.1: Comparison of FR and NR approaches for measurements at the point of the subscriber

	FR	NR
Technology	Direct comparison of Reference- and Processed- Signal	Analysis of given content without an explicit Reference
Measurement Type	Intrusive: Reference has to be available to measurement site	Non-Intrusive: No availability of Reference necessary
Real-time	Results delayed for clip length + evaluation time	Results delayed for min. buffering- and evaluation- time
Accuracy	High, but works only for known source signals	Medium (content dependent) due to unknown source signal
Limitations	High resource requirements (CPU and storage). Processed video can have a better quality than the noisy source video because of noise filters. Alignment errors are possible	May confuse certain artefact-like content with artefacts. Black videos received can produce high MOS scores although the source videos were not black
Implementation	Typically on workstation	Workstation or end terminal
System requirements	Enough CPU power and memory	Fast capture devices

A.2 Degradations and Metrics

A.2.0 General

Perceptual video quality metrics should be capable of identifying artefacts which can be intuitively understood by the average consumer of video. Furthermore, the characteristic degradation targeted by each metric should be unique. Finally, a comprehensive suite of metrics addressing the most common artefacts should be provided so that a combination of them can be used to reliably determine an overall quality rating, i.e. MOS.

A.2.1 Jerkiness

Jerkiness is a perceptual measure of motion that does not look smooth (in the extreme case a frozen picture). Transmission problems such as network congestion or packet loss are the primary causes of jerkiness. Jerkiness can also be introduced by the encoder dropping frames in an effort to achieve a given bit rate constraint. Finally, a low or varying frame rate can also create the perception of jerky motion. Jerkiness can be detected with the FR and the NR model.

A.2.2 Freezing

Video will play until the buffer empties if no new (error-checked/corrected) packet is received. If the video buffer empties, the video will pause (freeze) until a sufficient number of packets are buffered again. This means that in the case of heavy network congestion or bad radio conditions, video will pause without skipping during re-buffering, and no video frames will be lost. Freezing can be detected with the FR and the NR model.

A.2.3 Blockiness

Blockiness is a perceptual measure of the block structure that is common to all block-DCT based image and video compression techniques. The DCT is typically performed on 8x8 blocks in the frame, and the coefficients in each block are quantized separately, leading to discontinuities at the boundaries of adjacent blocks. Due to the regularity and extent of the resulting pattern, the blocking effect is easily noticeable. Encoding induced Blockiness can be detected with the FR and the NR model.

A.2.4 Slice Error

In many coding schemes (e.g. the MPEG family), each picture can contain one or more "slices". The number of slices will typically increase as the complexity of the image increases. Slices are used by the decoder to recover from data loss or corruption. Whenever an error is encountered in the data stream that corrupts one or more slices, the decoder will normally advance to the beginning of the next intact slice. Usually, slice errors will appear as black bars in the image, although the effect of slice errors is dependent on the error recovery mechanism deployed by decoders. Slice errors can be detected with the FR model.

A.2.5 Blurring

Blur is a perceptual measure of the loss of fine detail and the smearing of edges in the video. It is due to the attenuation of high frequencies by coarse quantization, which is applied in every lossy compression scheme. It can be further aggravated by filters, e.g. for deblocking or error concealment, which are used in most commercial decoders to reduce the noise or blockiness in the video. Another important source of blur is low-pass filtering (e.g. digital-to-analogue conversion or VHS tape recording). Blurring can be detected with the FR and the NR model.

A.2.6 Ringing

Ringing is a perceptual measure of ripples typically observed around high-contrast edges in otherwise smooth regions (the technical cause for this is referred to as Gibb's phenomenon). Ringing artefacts are very common in wavelet-based compression schemes such as JPEG2000, but also appear in DCT-based compression schemes such as MPEG and Motion-JPEG. Ringing can only be detected with the FR model.

A.2.7 Noise

Noise is a perceptual measure of high-frequency distortions in the form of spurious pixels. It is most noticeable in smooth regions and around edges (edge noise). This can arise from noisy recording equipment (analogue tape recordings are usually quite noisy), the compression process, where certain types of image content introduce noise-like artefacts, or from transmission errors, especially uncorrected bit errors. Noise can only be detected with the FR model.

A.2.8 Colourfulness

Colourfulness is a perceptual measure of the intensity or saturation of colours as well as the spread and distribution of individual colours in an image. The range and saturation of colours can suffer due to lossy compression or transmission. Colourfulness can be detected with the FR and the NR model.

A.2.9 MOS Prediction

When determining the quality of video sequences in subjective experiments, each observer gives a quality rating to every test video. The average of these ratings over all observers is called MOS. Both FR and NR metrics have to predict MOS, which can serve as estimators for overall video quality. MOS prediction can be done with the FR and the NR model.

A.2.10 Comparison of NR and FR regarding metrics and Degradations

Table A.2: Comparison of FR and NR regarding metrics and degradations

	FR	**NR**
Jerkiness	Yes	Yes
Freezing	Yes	Yes
Blockiness	Yes	Yes
Slice Error	Yes	No
Blurring	Yes	Yes
Ringing	Yes	No
Noise	Yes	No
Colourfulness	Yes	Yes
MOS prediction	Yes	Yes

Annex B: Bibliography

- ETSI TR 122 960: "Universal Mobile Telecommunications System (UMTS); Mobile multimedia services including mobile Intranet and Internet services (3G TR 22.960)".

History

Document history		
V1.1.1	August 2005	Publication
V1.2.1	June 2009	Publication
V1.3.1	July 2017	Publication