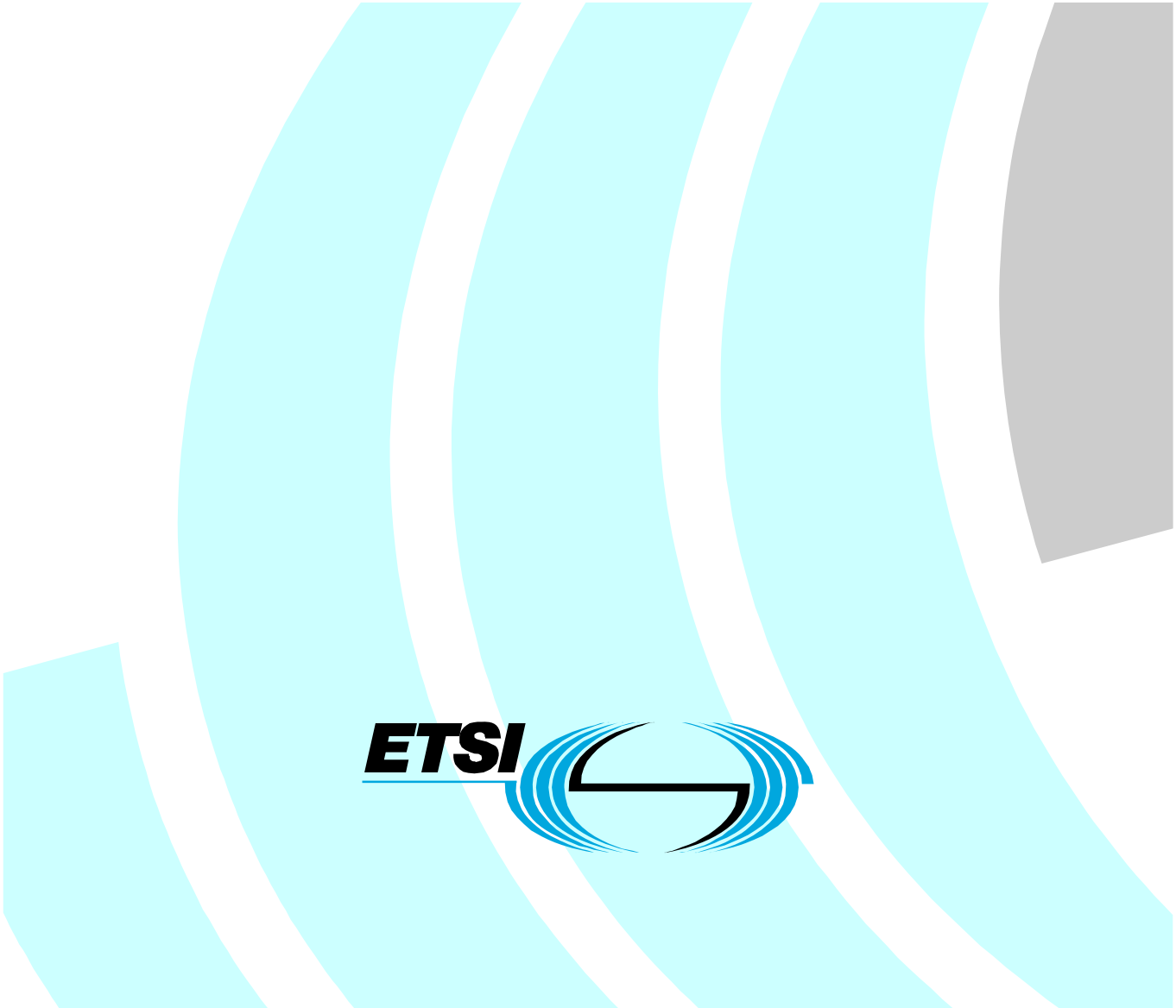
ETSI TR 102 493 V1.2.1 (2009-06)

Technical Report

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



Reference
RTR/STQ-00137m

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

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 (http://webapp.etsi.org/IPR/home.asp).

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.

Foreword

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

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

References are either specific (identified by date of publication and/or edition number or version number) or non-specific.

- For a specific reference, subsequent revisions do not apply.
- Non-specific reference may be made only to a complete document or a part thereof and only in the following cases:
  - if it is accepted that it will be possible to use all future changes of the referenced document for the purposes of the referring document;
  - for informative references.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

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

2.1 Normative references

The following referenced documents are indispensable for the application of the present document. For dated references, only the edition cited applies. For non-specific references, the latest edition of the referenced document (including any amendments) applies.

Not applicable.

2.2 Informative references

The following referenced documents are not essential to the use of the present document but they assist the user with regard to a particular subject area**.** For non-specific references, the latest version of the referenced document (including any amendments) applies.

[i.1] ETSI TS 126 233: "Universal Mobile Telecommunications System (UMTS); LTE; End-to-end transparent streaming service; General description (3GPP TS 26.233 version 8.0.0 Release 8)".

[i.2] VQEG: "Multimedia Group: Test Plan", Draft Version 1.5, March 2005.

[i.3] ETSI TS 122 960: "Universal Mobile Telecommunications System (UMTS); Mobile Multimedia services including mobile Intranet and Internet services".

[i.4] Final Report from the Video Quality Experts Group on the validation of the objective models of multimedia quality assessment, Phase.

[i.5] ITU-T Recommendation J.247: Objective perceptual multimedia video quality measurement in presence of a full reference.

[i.6] ITU-T Recommendation J.246: Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference.

[i.7] ETSI TS 126 114: "Universal Mobile Telecommunications System (UMTS); LTE; IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction (3GPP TS 26.114 Release 7)".

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:

BLER	BLock Error Rates
CIF	Common Intermediate Format (352 x 288 pixels)
DMOS	Difference Mean Opinion Score
FR	Full Reference Algorithm
HRC	Hypothetical Reference Circuit
ITU	International Telecom standardization Union
MOS	Mean Opinion Score
NR	No Reference Algorithm
PLR	Packet Loss Rates
PSNR	Peak Signal Noise Ratio
QCIF	Quarter Common Intermediate Format (176 x 144 pixesls)
RR	Reduced Reference
SRC	Source Reference Channel (or Circuit)
VGA	Video Graphics Adapter
VQEG	Video Quality Expert Group

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 perceived quality by the user using mobile terminal equipment. The goal is to have one or more objective video quality measurement algorithm(s), which predicts the video quality as perceived by a human viewer, which is in conformance with the minimum requirements list given in the present document.

On the input of the Video Quality Experts Group (VQEG) the ITU has recommended in ITU-T Recommendation J.247 [i.5] an objective perceptual video quality measurement in the presence of a full reference and in ITU-T Recommendation J.246 [i.6] a perceptual video quality measurement in the presence of a reduced reference. An objective perceptual multimedia video quality for no-reference algorithms has not been recommended. However continuing research within the VQEG is directed towards providing further input to the ITU on digital multimedia objective video quality measurement models. Work is going on in ITU-T and VQEG to develop and standardize hybrid, bitstream and parametric models.

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

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	Information loss
Video telephony	Two-way	32 kbps to 2 Mbps	< 150 ms preferred < 400 ms limit	< 80 ms	< 1 % pl
Streaming	One-way	32 kbps to 2 Mbps	< 10 s		< 1 % pl
Conversational Multimedia	Two-way		< 150 ms	Mutual service dependency, echo	

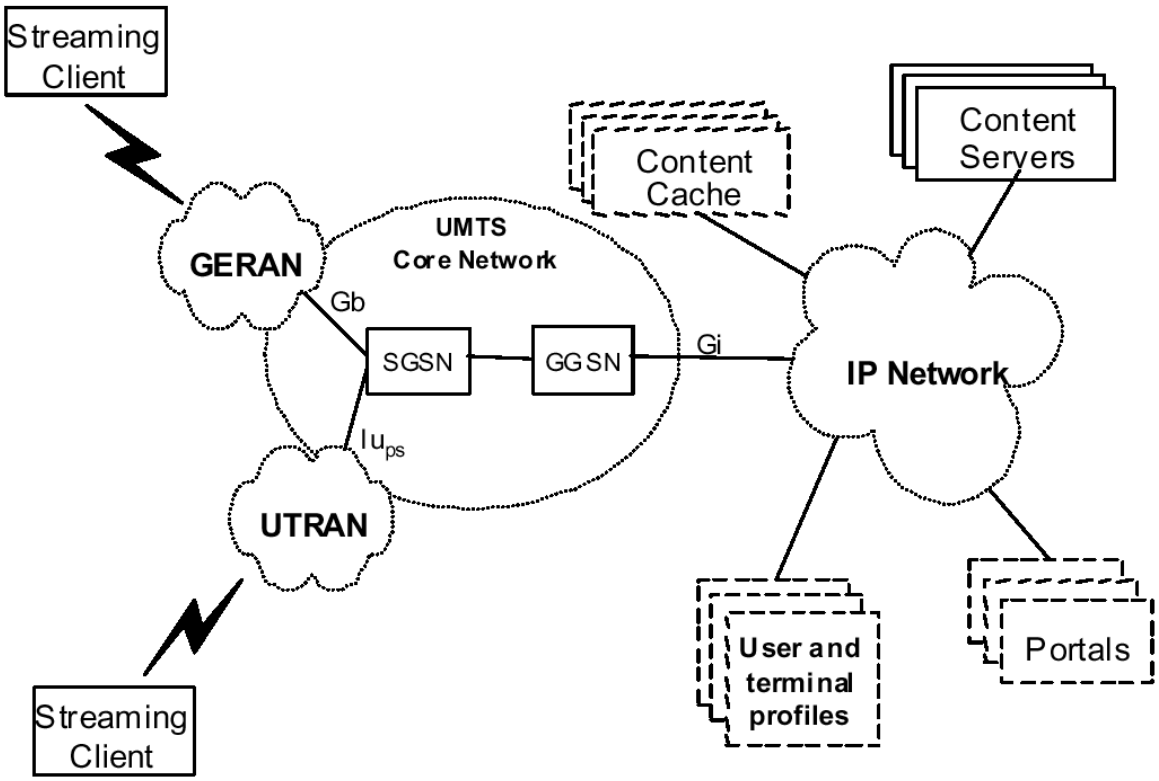


Figure 1: Streaming (TS 126 233 [i.1])

5.1 Streaming

Streaming refers to the ability of an application to play synchronized media streams like audio and video streams in a continuous way while those streams are being transmitted to the client over a data network. The client plays the incoming multimedia stream in real time as the data is received.

Typical applications can be classified into on-demand and live information delivery applications. Examples of the first group are music and news-on-demand applications. Live delivery of radio and television programs is an example of the second category.

For 3G systems, the 3G packet-switched streaming service (PSS) fills the gap between 3G MMS, e.g. downloading, and conversational services.

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. Microsoft Netmeeting is an example for a conversational multimedia application [i.3]. 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.7].

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

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 Key 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).
  - Bitstream (access to the transport layer payload, but not the video signal).
  - Hybrid (access to both the video signal and the transport layer payload).
  - Parametric (access to transport header, client information, and knowledge about used codecs).
- 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

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 itself 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) Sports.
4) Music video.
5) Advertisement.
6) Animation.
7) Broadcasting news (head and shoulders and outside broadcasting).
8) Home video.
9) Video Telephony (low quality input of various content).
10) Pictures /Still images.

Regarding 10) it is required that the algorithm can process pictures of the type of content delivered as moving picture (1 to 9) 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 Algorithm Properties

7.3.1 Full reference perceptual algorithms

In order to assure a wide range of applicability any full reference algorithm (FR) should be capable of working equally well with the uncompressed and a pre-processed (compressed) version of the reference. In cases where the reference is not loss less processed and hence the uncompressed original is not recoverable from the pre processed, an adequate mapping function has to be provided to facilitate homogeneous measurement results for both types of references.

For mobile environments the following scenario has to be taken into account:

An operator conveys live streaming as third party content to its users. In order to assess the end user quality of this content the capture on the end user side can only be compared with the stream as delivered by the content provider. If this is not being the uncompressed original but a processed one the operator needs to uncompress the delivery (see clause 7.3.3). This uncompressed stream serves as the reference for a FR assessment of the quality. If the compression was not loss less the 'original' is not recoverable and hence a FR algorithm applicable only for originals cannot be used.

7.3.2 No reference perceptual algorithms

No reference perceptual algorithms evaluate the quality without a dedicated undisturbed reference signal.

For perceptual no reference models erroneous evaluation is to be avoided. In particular that artefact-like content is not to be confused with real artefacts. Furthermore black videos received and freezing should not produce high MOS scores if the source of the videos was not black or a still image respectively.

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

7.3.3 No reference hybrid algorithms

A no reference hybrid algorithm uses both the video signal and the decoded bitstream to estimate the quality, but does not use the original video sequence as input for the quality estimation.

The algorithm is most likely able to distinguish between artefact-like content and video artefacts due to transport errors. The bitstream can be used to identify when the encoded data stream has been disturbed by transport errors.

A no reference hybrid algorithm can be used to score video for all scenarios.

7.3.4 Full reference hybrid algorithms

A full reference hybrid algorithm uses the original undisturbed sequence, the disturbed video signal and the received encoded bitstream to estimate the perceived quality.

The algorithm is able to distinguish artefact-like content and video artefacts due to transport errors. This type of algorithm is the one taking the largest amount if input to estimate the perceived quality.

A full reference hybrid algorithm can be used for all scenarios, but is not very well suited for live streaming.

7.3.5 Bitstream algorithms

A bitstream algorithm uses the encoded bitstream to estimate the perceived quality. It does not use the received video signal or the original video sequence.

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 consist of.

A bitstream algorithm is suitable for scenarios.

7.3.6 Parametric algorithms

A parametric media quality algorithm estimates the perceived quality based on measurement parameters, but not based on the video and audio signals themselves. 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 most 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.

7.3.7 Video Codecs

Video codecs may include but are not limited to:

- H.263.
- H.264.
- MPEG4.
- Real Video.
- SMPTE 421M VC-1 (Windows Media Video).

7.3.8 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 sequences.

7.4 Container schemes

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

- MPEG4.
- 3GPP.
- RM.
- AVI.
- FLV (Flash Video).

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 hence a value 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 Standardization of algorithms

The standardization of video quality algorithm has been divided into several phases from which one has been concluded. The MMI phase has been concluded with the publishing of two recommendations

8.1 Perceptual algorithms (J.246 and J.247)

On behalf of ITU-T Study Groups 9 and 12 and ITU-R Study Group 6 an informal group, called Video Quality Expert Group has conducted a study on perceptual video quality measurements. In this study [i.4], the Multimedia Phase I, perceptual video quality measurement algorithm has been evaluated. There were three types of algorithm for quality measurements in the presence of a full reference (FR), a reduced reference (RR) or no reference (NR). The evaluation was conducted on the formats QCIF, CIF and VGA.

The algorithm were tested on special material produced for this study and was controlled by extensive subjective testing. The test material has certain properties.

8.1.1 Sequence length

All original SRC source sequences were 12 seconds duration. The processed video sequences and the subjectively assessed reference files had duration of 8 seconds.

8.1.2 Content

The test material was selected from a common pool of video sequences.

The sequences have a frame rate of 25 or 30 frames per second

1) Video conferencing.
2) Movies, movie trailers.
3) Sports.
4) Music video.
5) Advertisement.
6) Animation.
7) Broadcasting news (head and shoulders and outside broadcasting).
8) Home video.
9) Video Telephony has no explicit tested however typical bandwidth and content is part of the content.

Still images were not among the material tested.

8.1.3 Formats

There were three different formats under test:

- PDA/Mobile (QCIF): (176 x 144 pixels).
- PC1 (CIF): (352 x 288 pixels).
- PC2 (VGA): (640 x 480 pixels).

NOTE: The newer format QVGA (320 x 240) was not tested. It can be conjectured that QVGA can be scored with CIF approved algorithms.

8.1.4 Bit Rates

The sequences for these formats had individually different bandwidths:

- PDA/Mobile (QCIF): 16 kbit/s to 320 kbit/s (e.g. 16, 32, 64, 128, 192, 320).
- PC1 (CIF): 64 kbit/s to 704 kbit/s (e.g. 64, 128, 192, 320, 448, 704).
- PC2 (VGA): 128kbit/s to 4Mbit/s (e.g. 128, 256, 320, 448, 704, ~1M, ~1.5M, ~2M, ~3M, ~4M).

8.1.5 Compressing algorithm

The coding schemes that could be used in that study included, but were not limited to:

- Windows Media Video 9.
- H.261.
- H.263.
- H.264 (MPEG-4 Part 10).
- Real Video (e.g. RV 10).
- MPEG1.
- MPEG2.
- MPEG4.
- JPEG 2000 Part 3.
- DiVX.
- H.264/MPEG4 SVC.
- Sorensen.
- Cinepak.
- VC1.

8.1.6 Container schemes

The used container scheme is audio video interleaved (avi). The sequences are considered as being taken from the display of a reproduction means (player). Processing and error concealment functions of a player are considered to be part of the Hypothetical Reference Circuit (HRC).

8.1.7 Evaluation

There has been a primary and a secondary evaluation of the algorithm. The primary analysis considers each video sequence separately while the secondary averages over different content under the same condition which reflects how the model tracks the average Hypothetical Reference Circuit (HRC) performance. The primary analysis is the most important determinant of a model's performance. Secondary analysis is presented to supplement the primary analysis.

8.1.8 Conclusions

In the result of the study five organizations have submitted an algorithm:

- NTT (Japan).
- OPTICOM (Germany).
- Psytechnics (UK).
- SwissQual (Switzerland).
- Yonsei University (Korea).

The VQEG formulated the following conclusions of the study has as result the following formulations:

Full Reference

- VQEG believes that some FR models perform well enough to be included in normative clauses of Recommendations. The scope of these Recommendations should be written carefully to ensure that the use of the models is defined appropriately.
- All four submitted FR models were close together in performance however in the primary evaluation Psytechnics and Opticom performed slightly better however not statistically significant in some cases as the models from NTT and Yonsei.

Reduced Reference

- VQEG believes that some of the RR models may be considered for standardization making sure that the scopes of these Recommendations are written carefully to ensure that the use of the models is defined appropriately.
- All four algorithm (with different side channel size) proposed by Yonsei University performed most of the time better than PSNR in the primary evaluation. The secondary analysis shows in principle a similar picture.

No Reference

- The VGA and CIF NR models did not perform well enough to be considered in normative portions of Recommendations.
- VQEG believes that the QCIF NR models may be considered for standardization making sure that the scopes of these Recommendations are written carefully to ensure that the use of the models is defined appropriately.
- In the secondary evaluation for QCIF the models proposed by Psytechnics and Swissqual performed occasionally better than the PSNR. The average correlations of the secondary analysis for the NR QCIF models were 0,91 for Psytechnics' model, 0,86 for SwissQual's model, and 0,81 for PSNR.

All three types were commented on as follows:

- None of the evaluated models reached the accuracy of the normative subjective testing.
- The secondary analysis requires averaging over a well defined set of sequences while the tested system including all processing steps for the video sequences have to remain exactly the same for all clips. Averaging over arbitrary sequences will lead to much worse results.

It should be noted that in case of new coding and transmission technologies, which were not included in this evaluation, the objective models can produce erroneous results. Here a subjective evaluation is required.

The results of the VQEG led the ITU to the publish two Recommendations J.246 [i.6] for reduced reference models and J.247 [i.7] for full referenced models.

8.2 Hybrid, bitstream and parametric algorithms

There are activities in ITU-T Study Group 9, ITU-T Study Group 12 and VQEG aiming at standardizing hybrid, bitstream and parametric algorithms. The activities are in currently (in 2009):

- VQEG is running a project with the goal to evaluate bitstream and hybrid algorithms.
- ITU-T Study Group 9 is the likely place for standardization of one or many hybrid algorithms.
- ITU-T Study Group 12 and Question 14 has two work items running: P.NBAMS which is a place-holder for a bitstream algorithm and P.NAMS a coming standard for a parametric audiovisual model.

Annex A (informative): Algorithms

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 algorithm that measures 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 algorithm 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

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*).

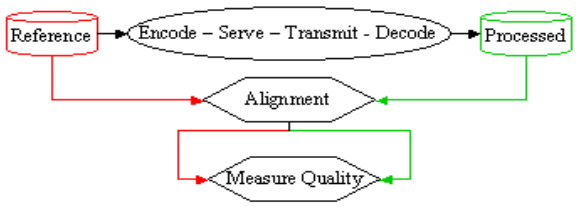


Figure A.1: Full Reference methodology

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.

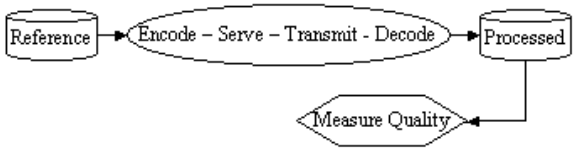


Figure A.2: No Reference methodology

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 'fooled' by content containing characteristics which look like an impairment.

EXAMPLES: 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.

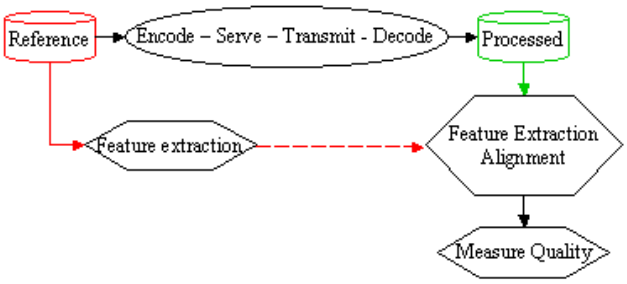


Figure A.3: Reduced Reference methodology

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 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.

Other models such as hybrid, bitstream, etc., models,

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

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

History

Document history		
V1.1.1	August 2005	Publication
V1.2.1	June 2009	Publication