ETSI TR 102 493 V1.1.1 (2005-08)

Technical Report

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

Reference

DTR/STQ-00081m

Keywords

algorithm, mobile, 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

Important notice

Copyright Notification

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 Processing, Transmission and Quality Aspects (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

For the purposes of this Technical Report (TR) the following references apply:

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

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

[3] ETSI TR 122 960: "Universal Mobile Telecommunications System (UMTS); Mobile multimedia services including mobile Intranet and Internet services (3G TR 22.960 version 3.0.1 Release 1999)".

[4] ITU-T Recommendation P.910: "Subjective video quality assessment methods for multimedia applications".

[5] ITU-R Recommendation BT.500-11: "Methodology for the subjective assessment of the quality of television pictures".

3 Definitions and abbreviations

3.1 Definitions

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

live streaming: streaming of live content

EXAMPLE: Web cam, TV programs, etc.

perceptual model: computational algorithm that aims to predict the subjectively perceived quality of video

streaming on demand: streaming of stored content

EXAMPLE: Movies.

3.2 Abbreviations

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

ACR	Absolute Category Rating
ADSL	Asymmetrical Digital Subscriber Line
CIF	Common Intermediate Format
CPU	Central Processor Unit
CRC	Cyclic Redundancy Check
DCT	Discrete Cosine Transform
DMOS	Difference Mean Opinion Score
DVD	Digital Versatile Disc
FR	Full Reference algorithm
GSM	General System for Mobile communications
HRC	Hypothetical Reference Circuit

HRR	Hidden Reference Removal
HSDPA	High Speed Downlink Packet Access
IP	Internet Protocol
ITU	International Telecom standardization Union
JPEG	Joint Photographic Expert Group
MOS	Mean Opinion Score
MPEG	Motion Picture Expert Group
NR	No Reference algorithm
PDA	Personal Digital Assistant
PLR	Packet Loss Ratio
PVS	Personal Video Station
QCIF	Quarter Common Intermediate Format
RR	Reduced Reference
VGA	Video Graphics Adapter
VHS	Video Home System
VQEG	Video Quality Expert Group
WCDMA	Wideband Code Division Multiple Access

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 gives 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.

At present there is no video quality algorithm standardized or approved by the ITU that meets those requirements. However continuing research within the Video Quality Experts Group (VQEG) is directed towards providing input to the ITU on digital multimedia objective video quality measurement models. The VQEG though advanced in their efforts did not give that input to the ITU yet (Any result is not expected before 3rd Quarter of 2005). Although the requirements for any eligible algorithm are outlined in detail there is no result so far. However the current requirements detail the form of the model, the focus for the multimedia-modelling component and the nature of the output necessary for the model to operate as a valuable assessment tool. Any algorithm proposed by the VQEG that will receive approval from the ITU will meet the requirements set by the VQEG therefore the present document will present those requirements.

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

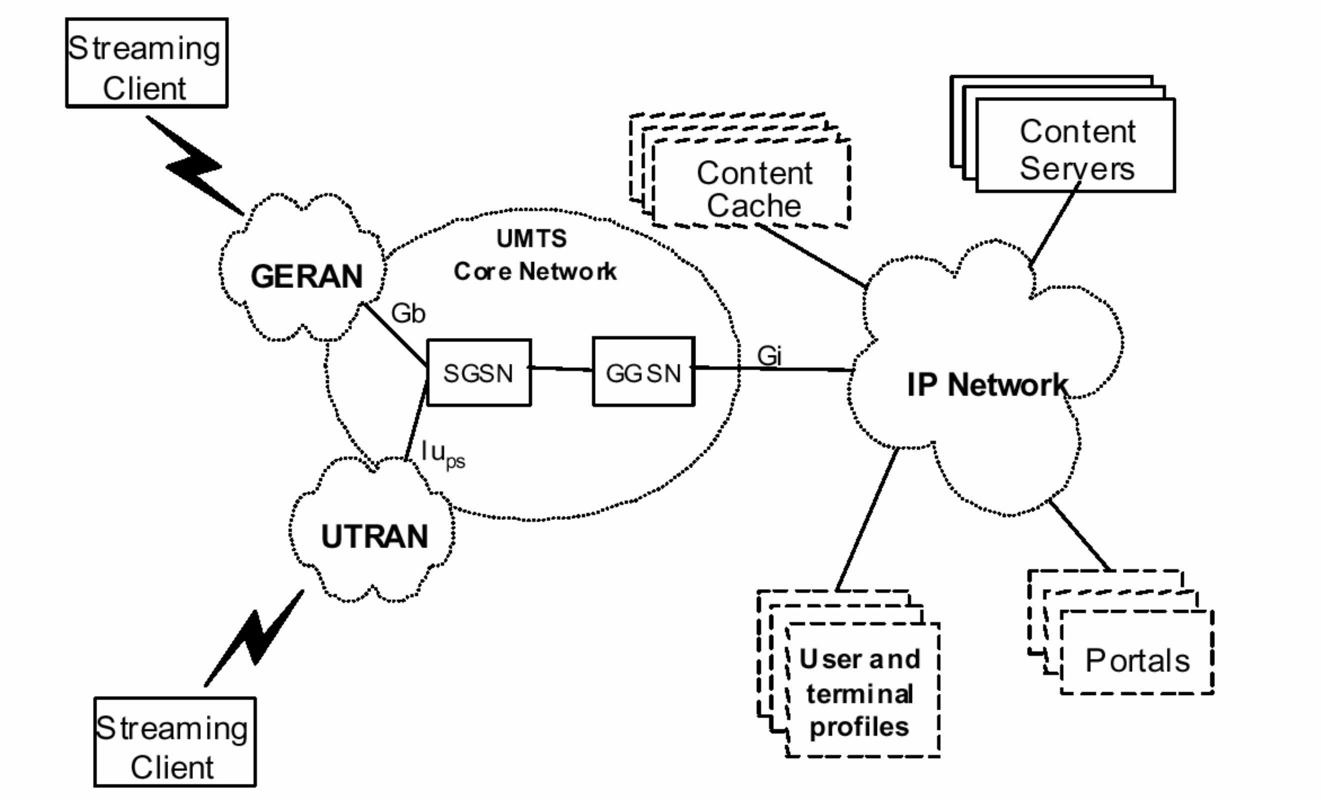

Figure 1: Streaming (TS 126 233 [1])

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 (TR 122 960 [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.

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 must 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

As the key scenarios there is live streaming, streaming on demand, video telephony and conversational multimedia. It can be tested as a drive test as much as it can be tested in a static fashion. The algorithm models that are used are Full Reference model (FR) and the Non Reference model (NR).

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

	FR	NR
Live Streaming	Yes	Yes
Streaming on Demand	Yes	Yes
Video Telephony	Yes	Yes
Conversational MM	Yes	Yes

Theoretically there are no restrictions to the applicability of the models for every key scenario. However for the FR model it is required that an appropriate reference is available (see clause 7.3.1).

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 both models. However due to the nature of the NR model it seems to be more suitable for that purpose. In the above scenarios low processing delay / immediate availability of the results outweigh the fact that NR models can be trapped e.g. with black frames etc.

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 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 to10 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. Therefore quality assessments that reflect the actual length and watching habits of viewers are desirable.

7.2 Content

The algorithm shall 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.

7.3 Algorithm Properties

7.3.1 Full reference 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 must be provided to facilitate homogeneous measurement results for both types of references.

For mobile environments the following scenario shall 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 algorithm

Erroneous evaluation is to be avoided in particular that artefact-like content is not confused with real artefacts. Furthermore black videos received shall not produce high MOS scores if the source of the videos was not black. This is for further study.

7.3.3 Compression algorithms

Compression algorithm may include but are not limited to:

- H.263;
- H.264;
- MPEG4;

- Real Video;
- Windows Media Video.

7.3.4 Calculation time

The calculation time should be as short as possible without any negative impact on the accuracy of the results.

7.4 Container schemes

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

- MPEG4.
- 3GPP.
- RM.
- AVI.

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 is standing for bad quality 5 for excellent quality.

Annex A: 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 must be 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*):

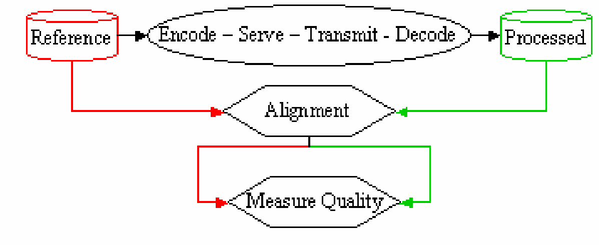

Figure A.1: Full Reference methodology

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 must 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:

1) Resources to store the Processed content must be made available.

2) 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.

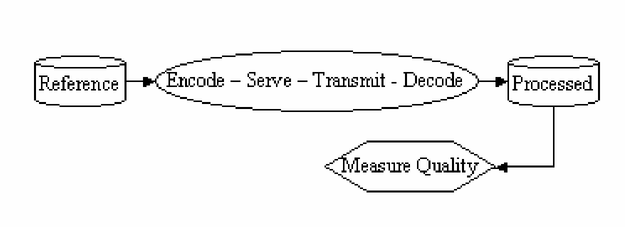

Figure A.2: No Reference Methodology

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.

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.

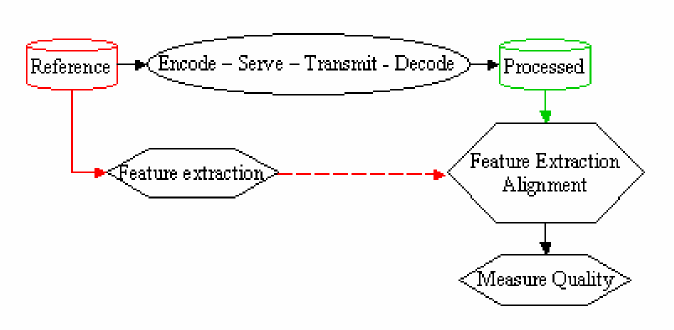

Figure A.3: Reduced Reference Methodology

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 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 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 must 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 is 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 8 x 8 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
Colorfulness	Yes	Yes
MOS prediction	Yes	Yes

Annex B: Characteristics required by VQEG

The VQEG has agreed on a draft test plan for the evaluation of video quality algorithm. The following clause refers to that test plan as the requirements for any video quality algorithm that desires to participate in that evaluation by the VQEG.

B.1 Formats

The algorithms should be able to handle at least one of the following video sizes. The defined sizes are:

- QCIF (176 × 144 pixels).
- CIF (352 × 288 pixels).
- VGA (640 × 480 pixels).

NOTE: 1 pixel of video will be displayed as 1 pixel native display. No up-sampling or down-sampling of the video is allowed at the player.

Presently, the algorithms have to perform video tests only. In the future also audio-video tests should be validated.

B.2 Types

Three different model types are allowed:

- Full Reference (FR).
- Reduced Reference (RR).
- No Reference (NR).

The side channels allowable for the RR models are:

- PDA/Mobile (QCIF): (1 k, 10 k).
- PC1 (CIF): (10 k, 64 k).
- PC2 (601): (10 k, 64 k, 128 k).

There can be one model (or algorithm) for each combination of model type and image size. For the RR model type there can be additionally a model for each side channel. Thus a total of 13 different models exist.

Table B.1: Algorithms - combinations of model types and side channels

	QCIF		CIF		VGA		
Full Reference							
Reduced Reference	1 k	10 k	10 k	64 k	10 k	64 k	128 k
No Reference							

Since the reduced reference model requires a side channel used during the main focus for mobile testing should be the Full Reference and No Reference.

B.3 Test materials

B.3.1 File Format

All source and processed video sequences will be stored in Uncompressed AVI in UYVY color space.

Source Frame Rate (SFR) is the absolute frame rate of the original source video sequences. The source frame rate is constant and may be either 25 fps or 30 fps.

Source material with a source frame rate of 29,97 fps will be manually assigned a source frame rate of 30 fps prior to being inserted into the common pool of video sequences.

B.3.2 Content

Typically the models should perform on a representative of a range of content and applications. The list below identifies the type of test material that forms the basis for selection of sequences.

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.

B.4 Degradations

Algorithms should be able to analyze different error conditions. These error conditions may include, but will not be limited to, the following:

- Compression errors (such as those introduced by varying bit-rate, codec type, frame rate and so on).
- Transmission errors.
- Post-processing effects.
- Live network conditions.

B.4.1 Simulated transmission errors

Simulated transmission errors are defined as errors imposed upon the digital video bit stream in a highly controlled environment. Examples include simulated packet loss rates and simulated bit errors. Parameters used to control simulated transmission errors are well defined.

A set of test conditions (HRC) will include error profiles and levels representative of video transmission over different types of transport bearers:

- Packet-switched transport (e.g. 2G or 3G mobile video streaming, PC-based wire line video streaming).
- Circuit-switched transport (e.g. mobile video-telephony).

Packet-switched transmission

HRCs will include packet loss with a range of Packet Loss Ratios (PLR) representative of typical real-life scenarios.

In **mobile video streaming**, we consider the following scenarios:

1) Arrival of packets is delayed due to re-transmission over the air. Re-transmission is requested either because packets are corrupted when being transmitted over the air, or because of network congestion on the fixed IP part. 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 until a sufficient number of packets is 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. This case is not implemented in the current test plan.

2) Arrival of packets is delayed, and the delay is too large: These packets are discarded by the video client.

NOTE 1: A radio link normally has *in-order delivery*, which means that if one packet is delayed the following packets will also be delayed.

NOTE 2: If the packet delay is too long, the radio network might drop the packet.

3) Very bad radio conditions: Massive packet loss occurs.

4) Handovers: Packet loss can be caused by handovers. Packets are lost in bursts and cause image artefacts.

NOTE 3: This is valid only for certain radio networks and radio links, like GSM or HSDPA in WCDMA. A dedicated radio channel in WCDMA uses soft handover, which not will cause any packet loss.

Typical radio network error conditions are:

- Packet delays between 100 ms and 5 seconds.

In **PC-based wire line video streaming**, network congestion causes packet loss during IP transmission.

In order to cover different scenarios, we consider the following models of packet loss:

- Bursty packet loss- The packet loss pattern can be generated by a link simulator by a bit or block error model, such as the Gilbert-Elliott model.
- Random packet loss
- Periodic packet loss.

NOTE 4: The bursty loss model is probably the most common scenario in a "normal" network operation. However, periodic or random packet loss can be caused by a faulty piece of equipment in the network. Bursty, random, and periodic packet loss models are available in commercially-available packet network emulators.

Choice of a specific PLR is not sufficient to characterize packet loss effects, as perceived quality will also be dependent on codecs, contents, packet loss distribution (profiles) and which types of video frames were hit by the loss of packets. For our tests, we will select different levels of loss ratio with different distribution profiles in order to produce test material that spreads over a wide range of video quality. To confirm that test files do cover a wide range of quality, the generated test files (i.e. decoded video after simulation of transmission error) will be:

1) Viewed by video experts to ensure that the visual degradations resulting from the simulated transmission error spread over a range of video quality over different contents.

2) Checked to ensure that degradations remain within the limits stated by the test plan (e.g. in the case where packet loss causes loss of complete frames, we will check that temporal misalignment remains with the limits stated by the test plan).

Circuit-switched transmission

HRCs will include bit errors and/or block errors with a range of bit error rates (BER) or/and block error rates (BLER) representative of typical real-world scenarios. In circuit-switched transmission, e.g. video-telephony, no re-transmission is used. Bit or block errors occur in bursts.

NOTE: Note that the term "block" does not refer to a visual degradation such as blocking errors (or blockiness) but refers to errors in the transport stream (transport blocks).

In order to cover different scenarios, the following error levels can be considered:

Air interface block error rates: Normal uplink and downlink: 0,3 %, normally not lower. High value uplink: 0,5 %, high downlink: 1,0 %. To make sure the proponents' algorithms will handle really bad conditions up to 2 % to 3 % block errors on the downlink can be used.

Bit stream errors: Block errors over the air will cause bits to not be received correctly over the air. A video telephony (H.223) bit stream will experience CRC errors and chunks of the bit stream will be lost.

B.4.2 Transmission Errors

Transmission errors are defined as any error imposed on the video transmission. Example types of errors include simulated transmission errors and live network conditions.

B.4.3 Live Network Conditions

Live Network Conditions are defined as errors imposed upon the digital video bit stream as a result of live network conditions. Examples error sources include packet loss due to heavy network traffic, increased delay due to transmission route changes, multi-path on a broadcast signal, and fingerprints on a DVD. Live network conditions tend to be unpredictable and unrepeatable.

Simulated errors are an excellent means to test the behaviour of a system under well defined conditions and to observe the effects of isolated distortions. In real live networks however usually a multitude of effects happen simultaneously when signals are transmitted, especially when radio interfaces are involved. Some effects like e.g. handovers can only be observed in live networks.

The term "live network" specifies conditions which make use of a real network for the signal transmission. This network is not exclusively used by the test setup. It does not mean that the recorded data themselves are taken from live traffic in the sense of passive network monitoring. The recordings may be generated by traditional intrusive test tools, but the network itself must not be simulated.

Live network conditions of interest include radio transmission (e.g. mobile applications) and fixed IP transmission (e.g. PC-based video streaming, PC to PC video-conferencing, best-effort IP-network with ADSL-access). Live network testing conditions are of particular value for conditions that cannot confidently be generated by network simulated transmission errors. Live network conditions should exhibit distortions representative of real-world situations that remain within the limits stated elsewhere in this test plan.

Normally most live network samples are of very good or best quality. To get a good proportion of sample quality levels, an even distribution of samples from high to low quality should be saved after a live network session.

NOTE: Keep in mind the characteristics of the radio network used in the test. Some networks will be able to keep a very good radio link quality until it suddenly drops. Other will make the quality to slowly degrade.

Samples with perfect quality do not need to be taken from live network conditions. They can instead be recorded from simulation tests.

Live network conditions as opposed to simulated errors are typically very uncontrolled by their nature. The distortion types that may appear are generally very unpredictable. However, they represent the most realistic conditions as observed by users of e.g. 3G networks.

Recording PVSs under live network conditions is generally a challenging task since a real hardware test setup is required. Ideally, the capture method should not introduce any further degradation.

For applications including radio transmissions, one possibility is to use a laptop with e.g. a built-in 3G network card and to download streams from a server through a radio network. Another possibility is the use of drive test tools and to simulate a video phone call while the car is driving. In order to simulate very bad radio coverage, the antenna may be wrapped with some aluminium foil (Editors note: This strictly a simulation again, but for the sake of simplicity it can be accepted since the simulated bad coverage is overlaid with the effects from the live network).

B.4.4 Video bit-rates

The algorithms should be able to perform on material encoded with the following bit-rates:

- PDA/Mobile: 16 kbs to 320 kbs (e.g. 16; 32; 64; 128; 192; 320).
- PC1 (CIF): 128 kbs to 704 kbs (e.g. 128; 192; 320; 448; 704).
- PC2 (VGA): 320 kbs to 4 Mbs (e.g. 320; 448; 704; ~1M; ~1,5M; ~2M; 3M;~4M).

B.4.5 Coding Schemes

Coding Schemes that will be used may include, but are not limited to:

- Windows Media Player 9.
- H.263.
- H.264 (MPEG-4 Part 10).
- Real Video (e.g. RV 10).
- MPEG 4.

B.4.6 Frame rates

For those codecs that only offer automatically set frame rate, this rate will be decided by the codec. Some codecs will have options to set the frame rate either automatically or manually. For those codecs that have options for manually setting the frame, 5 fps will be considered the minimum frame rate for VGA and CIF, and 2,5 fps for PDA/Mobile.

Manually set frame rates (new-frame refresh rate) may include:

- PDA/Mobile: 30; 25; 15; 12,5; 10; 8; 5; 2,5 fps.
- PC1 (CIF): 30; 25; 15; 12,5; 10; 8; 5 fps.
- PC2 (VGA): 30; 25; 15; 12,5; 10;8; 5 fps.

Temporally varying frame rates are acceptable.

B.5 Transmission Errors

Error conditions produced using packet loss rates and bit errors:

- Level 1: None.
- Level 2: Low.
- Level 3 Medium.
- Level 4: High.

B.5.1 Pre-Processing

The pre-processing may include, typically prior to the encoding, one or more of the following:

- Filtering.
- Simulation of non-ideal cameras (e.g. mobile).
- Colour space conversion (e.g. from 4:2:2 to 4:2:0).

B.5.2 Post-Processing

The following post-processing effects may be used:

- Colour space conversion.
- De-blocking.
- Decoder jitter.

B.5.3 Anomalous Frame Repetition

Anomalous frame repetition is defined as an event where the HRC outputs a single frame repeatedly in response to an unusual or out of the ordinary event. Anomalous frame skipping includes but is not limited to the following types of events: an error in the transmission channel, a change in the delay through the transmission channel, limited computer resources impacting the decoder's performance, and limited computer resources impacting the display of the video signal.

B.5.4 Pausing Without Skipping

Pausing without skipping (formerly frame freeze) is defined as any event where the video pauses for some period of time and then restarts without losing any video information. Hence, the temporal delay through the system must increase. One example of pausing without skipping is a computer simultaneously downloading and playing an AVI file, where heavy network traffic causes the player to pause briefly and then continue playing. A processed video sequence containing pausing without skipping events will always be longer in duration than the associated original video sequence.

Pausing without skipping events will not be included in the current testing.

B.5.5 Pausing with skipping

Pausing with skipping (formerly frame skipping) is defined as events where the video pauses for some period of time and then restarts with some loss of video information. In pausing with skipping, the temporal delay through the system will vary about an average system delay, sometimes increasing and sometimes decreasing. One example of pausing with skipping is a pair of IP Videophones, where heavy network traffic causes the IP Videophone display to freeze briefly; when the IP Videophone display continues, some content has been lost. Another example is a videoconferencing system that performs constant frame skipping or variable frame skipping. A processed video sequence containing pausing with skipping will be approximately the same duration as the associated original video sequence.

Pausing with skipping events will be included in the current testing. Anomalous frame repetition is not allowed during the first 1 s or the final 1 s of a video sequence. Note that where pausing with skipping and anomalous frame repetition is included in a test then source material containing still sections should form part of the testing.

If it is difficult or impossible to determine whether a video sequence contains pausing without skipping or pausing with skipping, the video sequence will be given the benefit of doubt and considered to contain pausing with skipping.

B.6 Registration

Full Reference Models must include registration.

Reduced-Reference Models must include temporal registration if the model needs it.

Temporal misalignment of no more than ±1 s or 10 % of the clip length (first occurrence will be taken) is allowed. Spatial offsets are expected to be very rare. Spatial registration will be assumed to be within (1) pixel. Gain, offset, and spatial registration will be corrected, if necessary, to satisfy the registration requirements specified.

No-Reference Models should not need registration.

B.6.1 Validation of algorithm

The selected test methodology for the subjective test is the single stimulus Absolute Category Rating method with hidden reference removal (henceforth referred to as ACR-HRR). This choice has been selected due to the fact that ACR provides a reliable and standardized method (ITU-R Recommendation BT.500-11 [5]; ITU-T Recommendation P.910 [4]) that allows a large number of test conditions to be assessed in any single test session. The reference will be included as one of the conditions. During the analysis the Hypothetical Reference Circuit (HRC) scores will be subtracted from the reference scores to obtain a DMOS score.

In the ACR test method, each test condition is presented once only for subjective assessment. The test presentation order is randomized according to standard procedures (e.g. Latin or Graeco-Latin square). At the end of each test presentation, subjects provide a quality rating using the ACR rating scale.

NOTE: For any reference-processed test condition pairing, a difference mean opinion score is calculated by subtracting the subjective rating for the processed condition from that of the reference condition.

History

Document history		
V1.1.1	August 2005	Publication