ETSI TR 102 714 V1.1.1 (2011-08)

Technical Report

Speech and multimedia Transmission Quality (STQ); Multimedia quality measurement; End-to-end quality measurement framework

Reference

DTR/STQ-00113

Keywords

multimedia, QoS

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://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.

Foreword

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

Introduction

The present document describes a framework for end-to-end objective quality measurement of multimedia services. The present document outlines a framework that specifies measurement options (e.g. parametric measures, perceptual models) at different points in the transmission chain. The framework defines measurement options for performing pre-transmission quality assurance, in-service network quality measurement, and for client-device quality monitoring.

1 Scope

The present document provides a framework for end-to-end quality measurement of multimedia services. The framework defines measurement options for performing pre-transmission quality assurance, in-service network quality measurement, and for client-device quality monitoring at different points in the transmission chain. The document is defined for services delivered on an IP network and is concerned solely with non-interactive services (e.g. IPTV, video streaming).

The present document is intended to provide industry and metric developers with a framework for implementing the appropriate quality metric at each point in the transmission chain. Recommendations are provided on the type and purpose of metric(s) for use at the head-end, network and end device (e.g. STB). Wherever possible, existing standardised metrics and measurement methods being appropriate for use in measurement tools in the transmission chain belonging to the defined framework are identified.

2 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 reference document (including any amendments) applies.

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 necessary for the application of the present document.

Not applicable.

2.2 Informative references

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] ITU-T Recommendation P.800 (1996): "Methods for subjective determination of transmission quality".

[i.2] ITU-R Recommendation BT.500-10 (2000): "Methodology for the subjective assessment of the quality of television pictures".

[i.3] ITU-R Recommendation BT.1788 (2006): "SAMVIQ - subjective assessment methodology for video quality".

[i.4] ITU-R Recommendation BS.1116-1 (1997): "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems".

[i.5] ITU-R Recommendation BS.1284 (1997): "Methods for the subjective assessment of sound quality".

[i.6] ITU-R Recommendation BS.1534 (2001): "Method for the subjective assessment of intermediate quality level of coding systems".

[i.7] ITU-R Recommendation BS.775-2 (2006): "Multichannel stereophonic sound system with and without accompanying picture".

[i.8] ITU-R Recommendation BS.1286 (2007): "Methods for the subjective assessment of audio systems with accompanying picture".

[i.9] ITU-T Recommendation P.910 (1999): "Subjective video quality assessment methods for multimedia applications".

[i.10] ITU-T Recommendation P.911 (1998): "Subjective audiovisual quality assessment methods for multimedia applications".

[i.11] ITU-T Recommendation P.920 (1996): "Interactive test methods for audiovisual communications".

[i.12] ITU-T Recommendation G.1070: "Opinion model for video-telephony applications".

[i.13] IETF RFC 4445: "A proposed Media Delivery Index (MDI)".

[i.14] ITU-T Recommendation P.800.1: "Mean Opinion Score (MOS) terminology".

[i.15] ITU-T Recommendation J.144: "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference".

[i.16] ETSI TS 183 063: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); IMS-based IPTV stage 3 specification".

[i.17] IETF RFC 2330: "Framework for IP Performance Metrics".

[i.18] 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.19] ITU-T Recommendation J.247: "Objective perceptual multimedia video quality measurement in the presence of a full reference".

[i.20] ITU-T Recommendation J.341: "Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference".

[i.21] ITU-T Recommendation G.1081: "Performance monitoring points for IPTV".

[i.22] ITU-R Recommendation BT.1359-1: "Relative timing of sound and vision for broadcasting".

[i.23] NTIA Technical Memorandum TM-10-472 (September 2010): "Relating Audio and Video quality using CIF Video".

NOTE: Available at http://www.its.bldrdoc.gov/pub/ntia-rpt/10-472/.

[i.24] ITU-R Recommendation BS.1387: "Method for Objective Measurements of Perceived Audio Quality".

[i.25] ISO/IEC 13818: "Information technology -- Generic coding of moving pictures and associated audio information (MPEG 2, 9 parts)".

[i.26] ITU-T Recommendation H.264: "Advanced video coding for generic audiovisual services".

[i.27] SMPTE 421M: "Television - VC-1 Compressed Video Bitstream Format and Decoding Process, 2006".

[i.28] ISO/IEC 14496: "Information technology -- Coding of audio-visual objects" (MPEG 4; currently in 11 parts).

[i.29] ETSI ES 202 667: "Speech and multimedia Transmission Quality (STQ); Audiovisual QoS for communication over IP networks".

[i.30] ETSI TR 102 720: "Speech and multimedia Transmission Quality (STQ); Delay variation on unshared access lines".

[i.31] IEEE 802.3: "Local Area Networks".

[i.32] IEEE 802.1Q: "Virtual Bridged Local Area Networks".

[i.33] IEEE 802.1p: "Traffic Class Expediting and Dynamic Multicast Filtering".

[i.34] IETF RFC 3471: "Generalized Multi-Protocol Label Switching (GMPLS) Signalling Functional Description".

[i.35] IETF RFC 2474: "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers".

[i.36] IETF RFC 2460: "Internet Protocol, Version 6 (IPv6) Specification".

[i.37] IETF RFC 3595: "Textual Conventions for IPv6 Flow Label".

[i.38] IETF RFC 3697: "IPv6 Flow Label Specification".

[i.39] IETF RFC 792: "Internet Control Message Protocol".

[i.40] IETF RFC 4443: "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification".

[i.41] IETF RFC 4969: "Stream Control Transmission Protocol".

[i.42] ITU-T Recommendation G.107: "The E-model: a computational model for use in transmission planning".

[i.43] IETF RFC 4585: "Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF)".

[i.44] IETF RFC 4588: "RTP Retransmission Payload Format".

[i.45] ITU-T Recommendation G.722: "7 kHz audio-coding within 64 kbit/s".

[i.46] ITU-T Recommendation G.729: "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)".

3 Abbreviations

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

ACR	Absolute Category Rating
AQ	Audio Quality
AVQ	Audio-Visual Quality
CIF	Common Intermediate Format
CQ	Conversation Quality
DF	Delay Factor
DPI	Deep Packet Inspection
DTT	Digital Terrestrial Television
FEC	Forward Error Correction
FR	Full Reference
HDTV	High Definition TeleVision
IP	Internet Protocol
IPTV	IP TeleVision
LQ	Listening Quality
MDI	Media Delivery Index
MLR	Media Loss Rate
MOS	Mean Opinion Scores
MOSe	estimated MOS
MOSo	objective MOS
MOSp	predicted MOS
MOSs	subjective MOS
MOVs	Model Output Variables
MSE	Mean Square Error
NR	No Reference
P2P	Peer-to-Peer

PEAQ	Perceptual Evaluation of Audio Quality
PSNR	Peak Signal to Noise Ratio
QCIF	Quarter Common Intermediate Format
QOE	Quality of Experience
QoS	Quality of Service
RR	Reduced Reference
SCTP	Stream Control Transmission Protocol
SDTV	Simple Definition TeleVision
STB	Set Top Box
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
VGA	Video Graphics Array
VLAN	Virtual Local Area Network
VoD	Video on Demand
VQEG	Video Quality Experts Group
WLAN	Wireless Local Area Network
xDSL	x Digital Subscriber Line
XOR	eXclusive OR (logical functionality)

4 Measurement methods and tools

4.1 Subjective vs. objective assessments

The measurement of end-to-end quality of multimedia services, as perceived by users, can be achieve through two types of measure: the instrumental measure ("objective" methods) are based on the use of measurement tools that catch and analyze the signal in or at the output of the network or use network parameters, to predict a quality score. The validation of these tools is based on the perceptive measure ("subjective" methods).

The general principle of the subjective methods consists in presenting audio and/or video sequences processed by the system under study, to testers who are asked to give their opinion on the quality of the presented sequences. The differences between methodologies mainly lie on the way of presenting sequences (sequence is presented alone, with the reference i.e. non distorted signal, or with other sequences), on the type of scale (scales with five, seven, ten categories, or continuous) and on the labels describing the scale.

The most used scale is the quality scale given in table 1.

Table 1: Levels of end-to end perceived quality

Quality	Score
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

This scale is used either as a simple category scale either combined to a continuous scale from 0 to 100, the five categories Excellent, Good, Fair, Poor, Bad then being used as anchor point. The average of the obtained scores gives a score generally called MOS for "Mean Opinion Score".

Most of the recommended subjective methodologies address one single modality (ITU-T Recommendation P.800 [i.1] and P8xx series for speech quality, ITU-R Recommendations BT.500-10 [i.2] and BT.1788 [i.3] for video quality, ITU-R Recommendations BS.1284 [i.5], BS.1534 [i.6] and BS.1116-1 [i.4] for audio quality). Some recommendations make suggestions for evaluating one modality in an audiovisual context or in the presence of an accompanying signal in the other modality (ITU-R Recommendation BS.775-2 [i.7]) for multichannel audio with accompanying picture, ITU-R Recommendation BS.1286 [i.8] for the testing of audio systems with accompanying image, ITU-T Recommendation P.910 [i.9] for evaluating the one-way overall video quality for multimedia applications such as videoconferencing). Only ITU-T Recommendations P.911 [i.10] and P.920 [i.11] apply to audio-visual subjective assessment, in a non-interactive context (P.911) or in an interactive one (P.920).

However, subjective tests are time consuming and expensive due to the need for a dedicated room as well as for human resources since around a thirty observers are needed to get a reliable precision on the mean opinion scores. Therefore, objective measurement tools are often preferred. The validity of audio and video quality metrics used in these tools is based on a good correlation with human judgements.

4.2 Measurement approach

There are three possible approaches to estimate the video quality:

- Full reference (FR) approach.
- Reduced reference (RR) approach.
- No reference (NR) approach.

The **FR** approach requires that the original video sequence is available. The output of the (coding) system is compared with the original video sequence and the system performance is estimated.

The **RR** approach requires that features extracted from the undistorted input (e.g. in the form of a reduced bandwidth signal) is available. The output of the (coding) system is compared with the reduced reference and the system performance is estimated.

The **NR** approach estimates the performance using the output of the (coding) system only.

For the time being only FR and RR approaches are standardised. These as well as other simple FR calculation algorithms are described in clause 4.6.

In video quality, the whole of standardized or pre-standardized predictive models which have a measurement approach by using producing the MOSp on quality scale. The MOSp metrics include integration models which directly depend on several subset metrics, expressing spatial and temporal artefacts. If one of these metrics is not reliable enough, the predicted MOSp score becomes false if compared to human judgment.

This issue appears more specifically on No Reference models and in realistic conditions, where operators want supervise TV live services by using reliable models with a minimum of usage limitations. Then, vendors have been promoting these past few years several separated QOE metrics rather than a global scoring MOSp indicator. These metrics were produced following a frame by frame analysis on a period of time ranging from 5 sec to 10 sec. The metrics produced address mainly the blockiness, freeze frames and black screen detections. Nevertheless, the evaluation of performances should be carried out through a false positive /false negative scoring as described in table 2. The false positive and false negative should be the lowest and the true positive and true negative should be the highest for a good algorithm that may be compared with a human judgment. It become important to standardize this new measurement approaches and evaluates the performances of predictive models with adequate subjective methodologies in standard process.

Table 2: Performance assessment of video artefacts detection algorithms

Performances of algorithms	Algorithm: artefact is present	Algorithm: artefact is not present
Perceived artefacts by the viewer	True positive	False negative
No perceived artefacts by the viewer	False positive	True negative

4.3 Classification of objective media quality assessment tools

The ability to measure the quality of speech, audio and multimedia services is important to service providers and network operators in ensuring that content is prepared, transmitted and received at an appropriate quality standard. For services such as VoIP, IPTV, video streaming and mobile TV, it is necessary to be assured and monitor quality in order to identify and resolve problems before they negatively impact on customer experience. A number of measurement tools are available for performing quality assessments at different points in the transmission chain. Objective quality measurement tools may be broadly classified into three categories:

- picture difference methods;
- parametric methods;

- perceptual methods.

The goal of these measurement methods is to provide quality indicators. Detailed descriptions for these methods may be found elsewhere. For present purposes, it should be noted that these methods have quite different computational and operational requirements. Furthermore, the output from these methods ranges from measuring noise present in the signal through to predicting subjective quality. The appropriate type of measurement method is dependent on the requirements defined by the user.

This framework document will begin by providing a basic nomenclature associated with each form of measurement method. A general measurement framework will be provided with a brief description of each stage in the measurement chain. For each measurement point identified in the framework, measurement options will be described.

4.4 MOS principles and values

Three basic measurement methods are commonly used to perform objective quality assessment. Picture difference methods, such as MSE and PSNR provide a pure measure of the difference between a source signal and its processed counterpart. PSNR is perhaps the most commonly applied measurement, being used in video encoders and for evaluating encoder performance.

Parametric methods have mostly been designed for usage as transmission planning tools. For speech, ITU-T Recommendation G.107 [i.42] provides users with a set of input parameters (e.g. packet loss rate, latency, codec type) that are used to compute a rating factor. This rating factor indicates the impact of varying network performance levels on the quality of a speech service. ITU-T Recommendation G.1070 [i.12] defines a similar parametric method designed transmission planning of video telephony services. The RFC 4445 [i.13] defines a parametric method (called the media delivery index) that may be used for in-service network quality monitoring of video services. ITU-T SG12 is also currently working on new models P.NAMS (parametric model for objective evaluation of multimedia quality on IP streams) and P.NBAMS (parametric model for objective evaluation of video only quality on IP streams based on examination of bitstream). None of the parametric methods have been specifically defined to measure perceptual quality, although some parametric methods do allow an estimated MOS to be computed. To calculate accurate MOS values perceptual models should be applied.

Perceptual models typically include some representation of the human visual system. Perceptual models are specifically designed to measure subjective quality. As such, perceptual models will provide a predicted MOS. Perceptual models are available for the measurement of speech, audio and video. ITU-T Recommendation J.144 [i.15] provides details of suitable perceptual quality measurement tools for broadcast television services.

ITU-T Recommendation P.800.1 [i.14] includes a nomenclature associated with measurements provided by different types of objective methods for measuring speech quality (listening quality (LQ) and conversation quality (CQ)). The nomenclature is defined as follows:

- MOSe (MOS-LQE and MOS-CQE) - estimated MOS produced by parametric models.
- MOSo (MOS-LQO and MOS-CQO) - predicted MOS produced by perceptual models.
- MOSs (MOS-LQS and MOS-CQS) - subjective MOS produced by human subjects.

In terms of accuracy and reliability, MOSs is the most accurate measure followed by MOSo. MOSe is the least accurate measure of subjective quality. It should be noted that standardised methods producing a MOSo have been independently validated against subjective scores. This is not true of methods outputting a MOSe. The present document recommends that the MOSe calculation is avoided as this is not independently validated against subjective scores.

For audio-video services such as video conferencing, internet video and television services, it is recommended that the ITU-T Recommendation P.800.1 [i.14] nomenclature is modified as follows:

- MOSe - estimated MOS produced by parametric models
- MOSp - predicted MOS produced by perceptual models
- MOSs - subjective MOS produced by human subjects

It is further recommended that the output of different methods is confined to their natural task. Therefore, for the purpose of the present document the following outputs are recommended:

- PSNR produces a dB; reduced or no reference PSNR methods produce an estimated dB (dB_e).
- Parametric methods output a quality rating or index defined as (QR or QI).
- Perceptual methods produce a MOSp.

4.5 Audio quality algorithm

In order to objectively measure the overall quality of an audio file (i.e. bandwidth 20 Hz to 20 kHz), 3 main families of tools (or algorithms) exist:

1) The first one is based on the comparison of a "human hearing model" of the degraded audio file with a "human hearing model" of the original audio file (un-degraded). This means that the original version is available.

   A method known as PEAQ (Perceptual Evaluation of Audio Quality) was standardized in 1999 (ITU-R Recommendation BS.1387 [i.24]). This algorithm is based on generally accepted psychoacoustic principles such as model of the peripheral human ear and cognitive components that take into account higher level processes underlying quality judgment.

   In general, it compares a signal that has been processed in some way with the corresponding time aligned original signal. Concurrent frames of the original and processed signals are each transformed to a basilar membrane representation and differences are analyzed further as a function of frequency and time by a cognitive model. The latter extracts perceptually relevant features which are used to compute a measure of quality. A number of intermediary model output variables (MOVs) are available. A selected set of MOVs is mapped to an objective quality grade.

   The present document is dedicated to mono and stereo audio files. Few years ago, the ITU launched a call for proposal to widen the standard to 5.1 audio files. 3 models were under competition and are now collaborating. Results should be known this year (2011).

NOTE 1: This algorithm is sold as a tool named OPERA by OPTICOM company. Many versions of this algorithm exist as different companies (such as France Telecom) have built PEAQ from the ITU standard.

2) The second family of algorithms is based on the comparison of parameters extracted/calculated from the original audio file with parameters extracted/calculated from the degraded version. This kind of algorithms is dedicated to broadcast applications when the original audio file is not available at the same time and place as its degraded version.

3) Algorithms from the third family do not make any comparison at all as they are based on detection of specific degradation in the degraded audio file. No algorithm has been standardized so far.

NOTE 2: In the past some companies have investigated this area and offered different algorithms. Most of them were dedicated to digital TV broadcasting and were working together with the same kind of algorithms applied to video (not separately). Some of them investigated internet TV applications.

4.6 Video quality algorithms

4.6.1 ITU-T Recommendation J.144

ITU-T Recommendation J.144 [i.15] provides guidelines on the selection of appropriate objective perceptual video quality measurement equipment designed for use in digital cable television applications when the full reference video signal is available. The validation test material did not contain channel errors. This Recommendation defines objective computational models that have been shown to be superior to peak signal to noise ratio (PSNR) as automatic measurement tools for assessing the quality of video.

The Recommendation includes four objective methods for the assessment of perceptual video.

Historically, the models in this recommendation have been the first standardized using a reference full video reference. Many organizations have been proposed their algorithms: Sarnoff/Tektronix, NTIA, CPqD, Tektronix/Sarnoff, NHK/Mitsubishi Electric Corp, KDDI, EPFL, NASA, KPN/Swisscom CT.

The algorithms have been assessed and recommended by VQEG group to the ITU for the high correlation between scores subjective and objectives scores. The study has been made for SD, VGA, CIF and QCIF display formats. In SDTV format the bitrate was in the range from 2 Mb/s to 12 Mb/s in MPEG-2 MP@ML [i.25] for digital and PAL and NTSC for analogue. The models have been validated without degradations in the transmission (e.g. packet loss in IP networks). The models have not been tested with H.264 [i.26] codecs.

The applications are limited to test a transmission chain and to evaluate codec performances in the laboratory context. But, with these two cases, the video reference should be injected into the models when the unimpaired reference video signal is readily available at the measurement point at the video head-end level. The algorithms were not assess if the video reference is corrupted or if one of the native video format is downscaling or upscaling (e.g. SDTV interleave format is transform to progressive scan to QVGA format). In some case, it could be more suitable to use the last ITU-T Recommendation J.341 [i.20] for HDTV applications.

The recommendation does not provide the performance of spatial en temporal re-alignment for each algorithm. In realistic case, the broadcasting delay can exceed more 10s between some video service providers. In the future, it could be required to give the limitation of the video resynchronization system. Or, the proponents of model should provide the performances video resynchronization system and the use cases.

4.6.2 ITU-T Recommendations J.246 and J.247

The Recommendations provide guidelines on the selection of appropriate objective perceptual video quality measurement methods when a reduced reference signal (ITU-T Recommendation J.246 [i.18]) or a full reference signal (ITU-T Recommendation J.247 [i.19]) is available. Applications identified in the scope of the Recommendations include:

- Internet multimedia streaming.
- Some forms of IPTV video payloads (VGA was the maximum resolution considered in the validation test).
- Mobile video streaming over telecommunications networks.

The ITU-T Recommendation J.246 [i.18] has been standardized in 2008, after the ITU-T Recommendation J 144 [i.15]. The Organization NTT, OPTICOM, Psytechnics, Yonsei has provided some efficient models. The models were evaluated for QCIF, CIF and VGA formats the range of bitrate was respectively from16 kbit/s to 320 kbit/s, 64 kbit/s-2 Mbit/s, 128 kbit/s-4 Mbit/s. The solutions have been tested with transmission errors by simulating IP packet lost. The recommendation can support up to 2 seconds of freeze image (with skipping). Several coding technologies have been used for testing H.264/AVC (MPEG-4 Part 10) [i.26], VC-1 [i.27], MPEG-4 Part 2 [i.28]. These models use reduced video reference video and degraded video signals in order to estimate the video quality. For practical aspect, this means that the reduced video reference should be delivery without any disturbance; otherwise the MOSp will be uncorrected.

The ITU-T Recommendation J.247 [i.19] has been standardized in 2008; the recommendation provides the guidelines to measure the objective perceptual video quality measurement methods when a full reference signal is available for QCIF, CIF and VGA formats. The Organizations NTT, OPTICOM, Psytechnics and Yonsei have provided some efficient models. The tests conditions for models validation were same the ITU-T Recommendation J.246 [i.18]. The usage limitations are the same as described for ITU-T Recommendation J.246 [i.18].

4.6.3 ITU-T Recommendation J.341

The ITU-T Recommendation J.341 [i.20] has been standardized in 2010; the recommendation provides the guidelines to measure the objective video quality measurement methods for HDTV formats. The potential usage is dedicated for monitoring the video quality at the source, for storage or transmission applications and lab testing of video system. Many models were tested by VQEG but only SwissQual model was retained for their very good correlation between predictive and subjective scores. The use case of monitoring the IPTV or DTT services for monitoring live channel is not explicitly mentioned. The model in this Recommendation was evaluated for HDTV applications and not on talking-head content typical of video-conferencing scenarios. Two coding technologies have been used for testing the model, H.264/AVC (MPEG-4 Part 10) [i.26] and MPEG-2 [i.25]. The objective perceptual video quality measurement methods can be used when a full reference signal is available in HDTV formats (1 920 points x 1 080 lines) and for progressive and interleave scan in 29,97 fps and 25 fps. The downsampling conversion has been made 1 080 p to 720 p. The usage limitations are the same as described for ITU-T Recommendation J.247 [i.19].

4.6.4 Error calculation

The number of error calculation appears mostly on the reduced and full reference algorithms. In video, the error calculation of MOSp happens when misalignment problems occur. The spatial misalignment occurs when offset, gain of luminance values are respectively upper than 20 % and 10 %. The models do not support also vertical and horizontal shift, cropping and re-scaling. The temporal misalignment can occur when the freeze frame, the black frame are too long. Generally, the algorithm is not enable to prevent the users when these defaults happen and the MOSp computed is very low even if the perceived video quality can be good.

One widely used error calculation approach is PSNR. It is defined as the ratio between the maximum possible power of a signal and the power of the corrupting noise of the processed signal. The mathematical definition of PSNR is:

$$PSNR = 10\,lg\left[\frac{R^2}{MSE}\right]$$

Where:

R is the maximum pixel value of the image.

MSE is defined below:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I_{i,j} - K_{i,j}\right]^2$$

Where:

I and *K* are the original image (*I*) and the distorted image (*K*).

Since they are easy to implement PSNR and MSE are frequently used to describe the performance of video coding systems.

4.7 Media Delivery Index (MDI)

The Media Delivery Index is described in RFC 4445 [i.13]. The MDI has two components:

- Delay Factor (DF).
- Media Loss Rate (MLR).

RFC 4445 [i.13] is an Informational RFC. The objective is to assess how well a network can transport video and to troubleshoot problems in networks whose performance has deteriorated due to reconfiguration or changing traffic loads. MDI uses this information to estimate buffer state and returns an index value that can be used to identify problems in the network. It is in no way a perceptual measurement method and is probably the coarsest method for determining network performance problems that affect user experience.

The MDI (Media Delivery Index) is very parametric approach that indicates video quality levels. The metric can help system operator to alarm if video default can occur. The MLR depends on virtual the buffer size setting up before measuring the indicator. The MLR does not take in account of the concealment error to calculate the potential visual impact. No correlation is carried between MDI and subjective scores by VQEG or another organization in order to know the realistic performance of the model.

4.8 Non-intrusive evaluation models of multimedia quality

ITU-T Recommendation SG 12 is working on two non-intrusive (i.e. NR) parametric evaluation models of multimedia quality, P.NAMS and P.NBAMS.

P.NAMS is a model based on IP protocol information over transport formats such as: RTP (over UDP), MPEG2-TS (over UDP or RTP/UDP), 3GPP-PSS (over RTP The model predicts MOS on the ACR scale for audio and video parts of a multimedia stream in multi-media mobile streaming and IPTV applications.

The primary applications for this model are monitoring of transmission quality for operations and maintenance purposes. The P.NAMS model may be deployed both in end-point locations and at mid-network monitoring points. The location of the model together with the location of the measurement probe determines the mode of operation.

It is expected that the work will be completed by end of 2011.

P.NBAMS is a model based on IP protocol and bit-stream information. The model predicts the impact of observed coding and IP network impairments on quality experienced by the end-user in multi-media mobile streaming and IPTV.

The model can use the bitstream information in addition to information contained in packet headers, prior knowledge about the media stream, and buffering information from the client.

The model has two modes. In the first mode the model does not completely decode the payload. This means that no pixel information is used internally in the model. In the second mode the model decodes parts or all of the video sequence and the pixel information are used for the MOS estimation, in addition to other bitstream information.

The first version of this model predicts Mean Opinion Scores (MOS) on the five point ACR scale for the video part of the stream. The audio and multi-media MOS score may be considered in a second phase of P.NBAMS (Using the audio and multimedia part of P.NAMS is another option).

The primary applications for this model are monitoring of transmission quality for operations and maintenance purposes. The P.NBAMS model may be deployed both in end-point locations and at mid-network monitoring points.

The primary quality prediction made by such a model is based on the payload of the stream being analyzed. Therefore, this Recommendation can provide a comprehensive evaluation of quality as perceived by a particular end-user because its scores can reflect the impairments on the coding and the IP network being measured, which differ from user to user.

It is expected that the work will be completed early 2012.

4.9 Audio-Video Interaction

Previous clauses present audio and video quality algorithms which assume only single modality. However multimedia services imply both modalities together. Human perception system is highly non-linear. Therefore, the perception of an audiovisual event is not a simple linear combination or audio and visual perceptions done somewhere in the brain. In effect, there are audio-video interactions which may have physiological origins with the bimodal behaviour of nerve cells localised in the superior colliculus. Audio-video interactions are multiple and mainly depend on physical relationships between auditory and visual stimuli. But they also depend on the context of stimuli presentation, on their intrinsic and related meaning, or on the a priori knowledge of observers and their attention on stimuli.

In ecological situations, audio-video interactions are expressed by a natural sensorial integration resulting in perceptual fusion: sound and image form a unique coherent percept for the observer. To some extent, the perceptual fusion stands up to incoherence between audio and video flows. For instance, everyone in front of his/her television may experience the ventriloquism effect: the sound seems to come from the mouth of people who animate on screen where as it comes from two loudspeakers on both sides of the television.

There are two aspects addressed in this clause:

- Lip synchronization (Audio-Video synchronization).
- Audio-Video quality interaction.

4.9.1 Lip synchronization

One important issue of the overall audio-visual quality lies on the synchronisation between audio and video, since the perceived quality degrades rapidly when asynchrony is increased. Subjective evaluations show that detectability thresholds are about +45 ms to -125 ms and acceptability thresholds are about +90 ms to -185 ms on the average, a positive value indicates that sound is advanced with respect to vision [i.22]. However, the perception of asynchrony highly depends on the type of content and on the task.

For further information ITU-R Recommendation BT.1359-1 [i.22] should be consulted.

4.9.2 Quality interaction

For spatiotemporally coherent audiovisual systems, many studies show that when subjects are asked to judge the audio quality (AQ) of an audiovisual stimulus, the video quality (VQ) will contribute significantly to the perceived audio quality. To a lesser extent, when subjects are asked to judge the video quality of an audiovisual stimulus, the audio quality will contribute to the perceived video quality. In addition, there is a significant mutual influence between the auditory and the visual domain in the perceived overall audio-visual quality (AVQ). Most studies have shown that video quality dominates audiovisual quality in general, but audio quality would be more important for talking head content because the human attention is mainly focused on the auditory stimulus. Some studies have been done to derive AVQ based on single AQ and VQ. A commonly used fusion model proposes that the overall audio-visual quality can be derived by a linear combination and a multiplication of AQ and VQ where the multiplication of AQ and VQ has very high correlation with the overall quality. NTIA Technical Memorandum TM-10-472 [i.23] presents an overview of some models for calculating the overall audiovisual quality. However, for the time being, there is no reliable metric available for measuring the audio-visual quality automatically, since the relationship between AQ, VQ, and AVQ is also influenced by some other factors, such as attention of subjects, audiovisual content, usage context, and the experiment environment. One of the shortcoming of the present models is that no tests have been made using HD video.

4.10 Transmission degradation

4.10.1 Packet loss

Media (e.g. speech, audio or video) are usually transported over IP networks using an unreliable protocol which does not guarantee that packets are delivered or delivered in order. Packets may be dropped under peak loads and during periods of congestion (caused, for example, by link failures or inadequate capacity). Packets may also be dropped at the receiving device due to jitter buffer overflow.

Transmission interruption can be seen as packet loss.

4.10.2 Delay and delay variation

The delay sources of an IP network connection are:

- Transmitting terminal delay.
- Access network delay.
- Core network delay.
- Receiving terminal delay.

ES 202 667 [i.29] contains an overview of delay related to different parts of a connection and the technology used.

Transmission delay variations (jitter) are caused by queuing in network elements or by routing the packets along different network paths. A significant multimedia delay variation source is the simultaneous transmission of packets containing information for two or more media; e.g. voice and video. In serial access networks such as xDSL systems only one packet can be transmitted at an instant, other packets has to be delayed until the transmission link is free. TR 102 720 [i.30] provides an introduction on the effect of different IP services on lines with limited bandwidth (e.g. DSL). It explains the mechanism of serialisation delay gives an overview over prioritisation and shows how the maximum delay variation due to concurrent traffic can be calculated.

In wireless systems (e.g. WLAN or cellular systems) several connections share the wireless link. To get access to the wireless link a queuing mechanism is implemented. This mechanism increases the delay variation. The number of participants sharing the wireless link influences the amount of delay variation.

4.10.3 QoS mechanisms

When discussing QoS mechanisms, one should be aware of the multi -layered view that constitutes modern packed-based networks. Just as each layer has a smaller or larger degree of effect on the observed quality, related mechanisms can be employed to minimise the negative impact.

4.10.3.1 QoS mechanisms at link, network or transport layer

For the physical/link layer there are mechanisms related to each specific technology. Networks based on xDSL, Coaxial cable, radio, Twisted Pair (Ethernet) or fibre will all have one or several QoS mechanisms that are tailored to its operating environment. Some examples are:

- xDSL (buffer prioritizing)
- Ethernet [i.31]: the use of IEEE 802.1Q [i.32] tags in the Ethernet header frame to identify VLAN membership and IEEE 802.1p [i.33] to set priority
- Fibre (GMPLS [i.34]): provides QoS by traffic engineering based on labels, through switching in time, packet, wavelength or fibre domains

The network layer with its associated protocols plays an important role when it comes to controlling and minimizing transmission degradations. Even though both IPv4 and IPv6 are best effort by nature and can be characterized as unreliable, they have some built-in functions that relate to QoS:

- Differentiated Services Code Point (DSCP) [i.35] and Explicit Congestion Notification (ECN) in IPv4
- Traffic Class (DSCP + ECN) and Flow Label in IPv6 [i.36]

Use of DSCP and ECN are dependent on support from the network doing the actual transfer of IP packets. Flow Label specifications and minimum requirements are described in RFC 3595 [i.37] and RFC 3697 [i.38], and some early implementations are underway.

ICMPv4 [i.39]/ICMPv6 [i.40] also belong to the network layer and are among the core protocols of the Internet Protocol suite. Both protocols assist network nodes by exchanging error messages that indicate the nature of a broad selection of fault conditions, e.g. congestions.

Protocols at the transport layer hold some of the most important QoS mechanisms used in IP networks. This is especially true for TCP, which together with IP and DNS forms the basis for Internet. World Wide Web, E-mail, Secure Shell (remote administration), File Transfer Protocol, P2P and some streaming media applications are services that rely on TCP and its reliable delivery function, assisted by the following mechanisms:

- Flag fields (control bits), used by nodes to express capabilities related to network and payload handling
- Window size, for implementation of flow control in an environment where nodes of different capabilities communicate
- Congestion control, controlling the rate of data entering the network so as to avoid trigging collapse scenarios

UDP is another transport protocol that provides datagram service. It is often used by applications that prefer reduced latency over reliability. Even though UDP lack any form of loss, flow or congestion control, it is still an important component in network quality as it provides real-time applications with an effective transport service.

Stream Control Transmission Protocol (SCTP) is defined by IETF in RFC 4969 [i.41]. SCTP provides some of the same capabilities as both TCP and UDP in that it is message-oriented (like UDP) but at the same time provides reliable, sequence controlled and congestion controlled transport services (like TCP). However, some applications need reliable transfer but have little need for a strict sequence maintenance. Others need only partial ordering of data - in both these instances, TCP would cause unwanted latency and SCTP will thus be better suited to the task.

4.10.3.2 QoS mechanisms at application layer

With the mechanisms used for video compression, IP packet loss has a significant impact on the quality of the image. The loss of a single IP packet can cause a temporary "Pixelation" or "Freeze" screen. Packet loss of 100,000, which is quite acceptable for most IP applications, causes a screen degradation every 5 minutes which is unacceptable for video application (IPTV).

In this condition, the idea is to decrease packet loss sensitivity in IPTV service by using application-level mechanisms: the Forward Error Correction (FEC) or retransmission of lost packets.

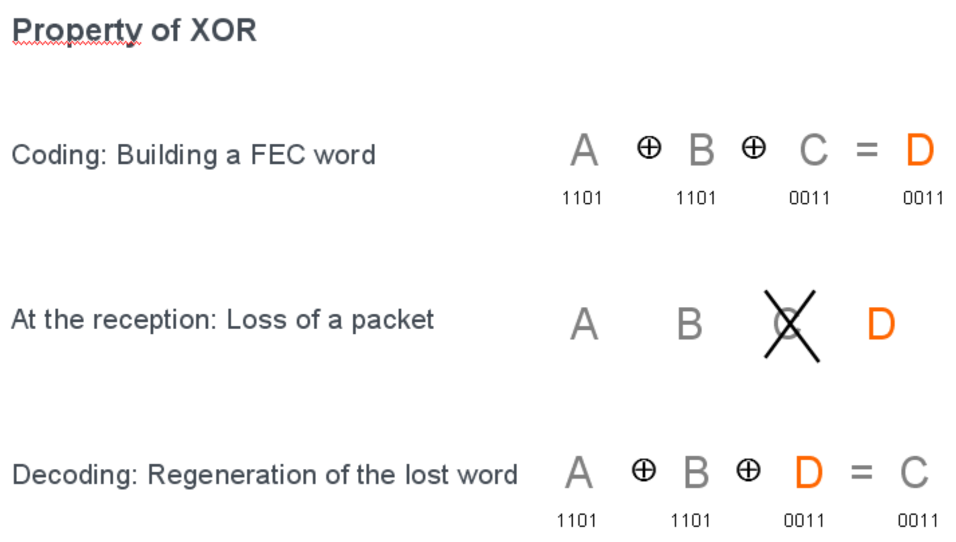
Generic principle of a FEC mechanism (Forward Error Correction) is to add some redundant information to the information to be sent in order to re-build missing information at the receiver side, using the redundant information.

Forward Error Correction (FEC) is a method of protecting data streams across networks where packet loss and delay are known to exist. Prior to transmission, the data is put through a pre-defined algorithm that adds extra FEC packets specifically for error correction at the receiving end. If a video packet is received in error, the FEC packets are used to check and repair the video so that the exact original packets can be passed on to the decoder. In this way, FEC acts as a shield that protects the video signal from the impairments imposed by an IP network, such as packet loss, packet delay and packet jitter.

This FEC can be sized for typical loss profiles: impulsive noises usually translate into burst packet losses, the size of the matrix will depend on the noise measured on the network. The FEC is equally well applicable on unicast or multicast streams and does not need any upstream capacity. In terms of scalability, the required bandwidth for FEC only depends on the number of protected channels. As depicted, impacts on architecture are rather low: for an end-to-end protection, they are mainly located at the head-ends where FEC inserters are to be added and at the set-top-boxes where FEC decoders are to be implemented.

The Pro-MPEG Forum has approved an open standard, Code of Practice #3 (COP #3), to address the issues of transporting video over IP networks. One problematic IP network characteristic, burst packet loss, is caused by buffer and re-route issues. COP #3 FEC can protect a video stream from a burst packet loss of up to 255 packets, which is suitable for most private, managed IP networks using QoS techniques such as MPLS, RSVP, and DiffServ.

The generation of FEC packets in the COP #3 standards is based upon a matrix defined by the parameters L and D. L represents the number of columns in the matrix, while D represents the number of rows. The standard defines the generation of two types of FEC packet: Column FEC and Row FEC. A FEC packet is generated by XOR of the media packets in a column or a row. Once generated, the Column FEC packets and Row FEC packets are transmitted along with the original media packets on 3 separate UDP ports to a Pro-MPEG COP #3 compliant receiving device.



Packet loss retransmission is other mechanisms to decrease packet loss sensitivity in IPTV service.

Standardized retransmission mechanism are presented in RFC 4585 [i.43] (RTCP feedback) and RFC 4588 [i.44] (RTP retransmission).

Basic principles of retransmission mechanism are:

- video stream is encapsulated in RTP (sequence number);
- when the STB detects a packet loss, it sends a RTCP_NACK message describing packets to retransmit.

An additional buffer is required in the STB, in order to delay the video display, allowing waiting any retransmitted packets.

4.11 IP protocols

Depending on what kind of services an application requests for transmitting content between parties, one or more protocols are selected. In addition to TCP and UDP, other protocols specially suited for transport of media content can be used. These include HTTP, SCTP and RTP.

TS 183 063 [i.16] specifies the protocol aspects of an IMS-based IPTV system. In the document it is stated that "The Media Distribution Function (MDF)" send the content using one of the following transport technologies:

- MPEG2 TS encapsulation.
- Direct RTP transport.

5 Quality Measurement Framework

A quality measurement framework is now introduced. The framework is confined to quality measurement methods that may be defined for transmission planning, network performance measurement (e.g. IPPM as set out in RFC 2330 [i.17]), quality assurance and monitoring obtained pre-transmission, during transmission and post-transmission. Quality of experience metrics (e.g. channel change times, EPG response functionality and ease of use, device-dependent customer expectations and so on) are for future study. Figure 1 provides a general overview of the quality measurement framework. Figure 1 identifies useful measurement points as well as appropriate types of measurement.

Something related to ITU-T Recommendation G.1081 [i.21]. Explain why a three stage model is used.

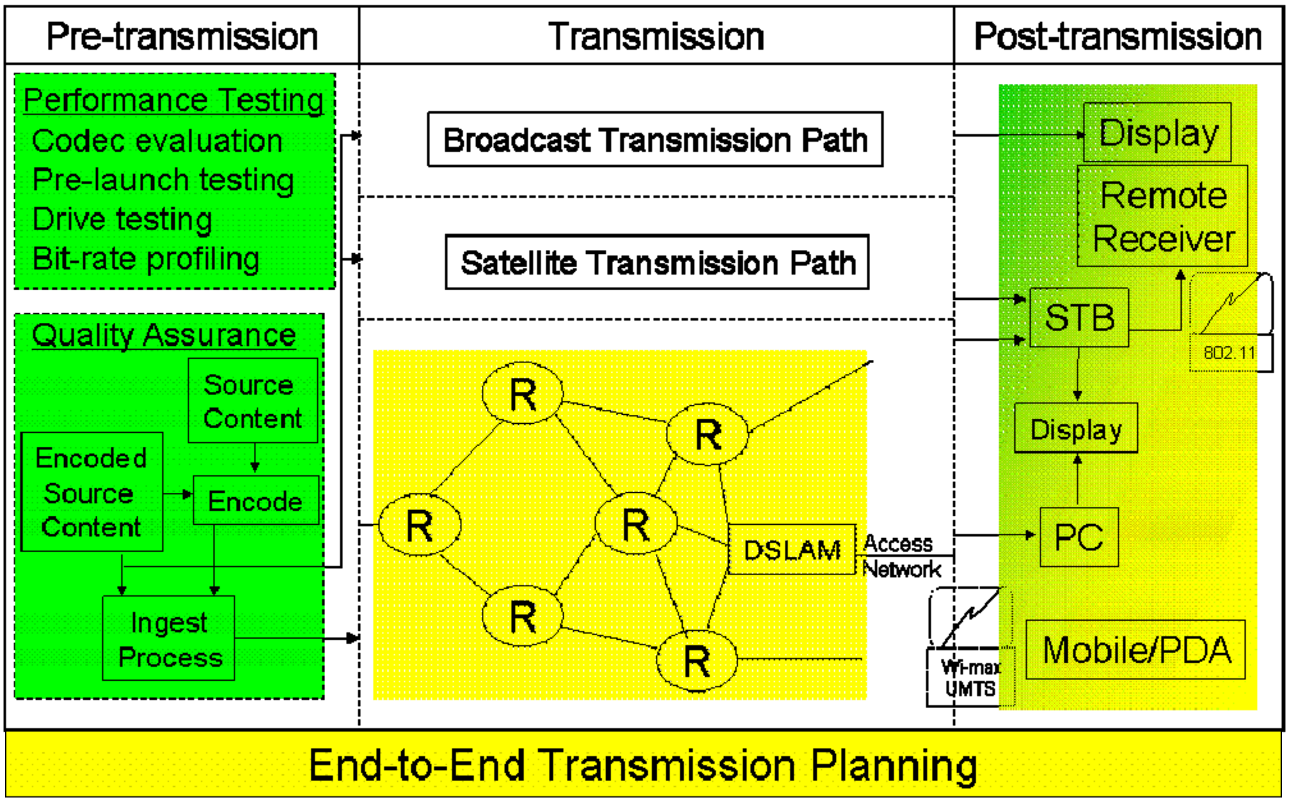


Figure 1: Quality Measurement Framework

In figure 1 Green regions indicate suitability for perceptual quality measurement. Yellow regions indicate parametric measurement as most suitable measure. Mixed green/yellow indicates both perceptual and parametric measurements may be usefully applied.

The framework can be delineated into three specific points of measurement. Firstly, measurements are obtained prior to transmission. The second point of measurement is in the network itself. Finally, measurements are obtained at the receiving device. Each measurement point will now be discussed in more detail.

5.1 Pre-transmission Measurement

Measurements derived pre-transmission may be used for a number of purposes, including performance testing and quality assurance. For performance testing, application developers can test the quality of media applications (e.g. codec development, error concealment methods). For industry, reliable measurement methods are necessary to evaluate the performance of competing vendor products (e.g. codec evaluation).

Once a service provider has decided upon their operational environment, quality measurements become important to check that content is received, encoded and ingested at appropriate quality prior to distribution to customers. Figure 2 illustrates the pre-transmission content processing chain. Content is often provided by third parties to service providers. The service provider may receive content at distribution quality (e.g. digibeta tapes) or already encoded for transmission. If service providers receive distribution quality content, they will need to encode the content themselves prior to transmission. Once content has been encoded according to the desired profile, an ingestion procedure is performed to place the media into the appropriate transmission package.

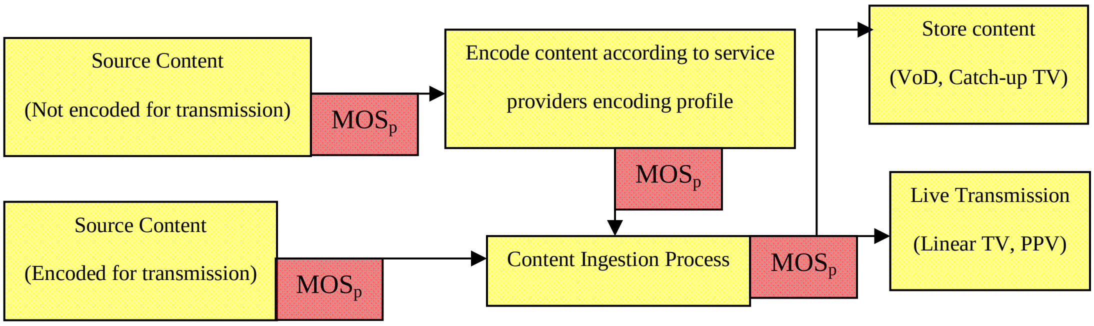


Figure 2: Pre-transmission quality assurance

For both performance testing and quality assurance, perceptual quality measurements are appropriate. Perceptual quality measurements, such as pixel-based methods using full reference procedures (ITU-T Recommendation J.144 [i.15]) are particularly useful for performance testing where different coding schemes (e.g. G.722 [i.45], G.729 [i.46], AAC+, MPEG-2 audio, MPEG-2 video, H.264, VC-1) are under evaluation. Full reference methods tend to be computationally intensive, require access to both the original and processed version of content and require spatio-temporal registration to be performed. As a result of these limitations, full reference methods are not ideal for quality assurance tasks. For quality assurance, real-time measurement methods are necessary. Furthermore, it is desirable that QA methods apply no reference procedures as the operational environment may not always allow access to the source content.

5.2 Network Measurement

There exist a number of methods for obtaining network performance measurements, the simplest of which reside on network routers that may be probed to determine traffic management efficiency and identify and quantify problems. More specific network measurement probes have been defined (e.g. RFC 4445 [i.13], RFC 2330 [i.17]) yet have failed to become commonplace due to the increased overhead and installation associated with their application. Perceptual quality metrics, particularly reduced reference and no reference methods have been identified as potential tools for in-service network based monitoring. Although all forms of measurement probe may be applied to network monitoring, simplicity is especially important for network measurement and it is expected that minimalist network measurement will be performed. As a result, existing channel statistics will remain as the mainstay of network measurement.

Deep Packet Inspection

Deep Packet Inspection (DPI) is a technology that is capable of inspecting in real-time headers and payload of every packet that passes through the DPI device, that means packet headers, types of applications, and actual packet content. DPI has the ability to inspect traffic at layers 2 through 7. Using this technology it is possible to classify individual flows on per-application or per-user basis.

The technology is controversial for security and privacy reasons. However, several products are available, and used by network operators to assess and manage network traffic; applications that fits well are out of the scope of the present document.

5.3 Post-transmission Measurement

The point of reception is the single most important measurement for monitoring customer experience. It may be argued that knowledge of quality of encoded, ingested and transmitted signal together with information relating to network performance (e.g. packet loss rate, latency) is sufficient to understand the user experience. Where this measurement scenario fails is that it does not consider problems with the receiving device (or home network environment) nor is it able to adequately accommodate error concealment. For accurate post-transmission measurement, no reference perceptual quality methods are most appropriate. Reduced reference methods may be applied where a sidechannel is made available; however the requirement for a sidechannel limits the usefulness of reduced reference methods. Figure 3 outlines post-transmission measurement options.

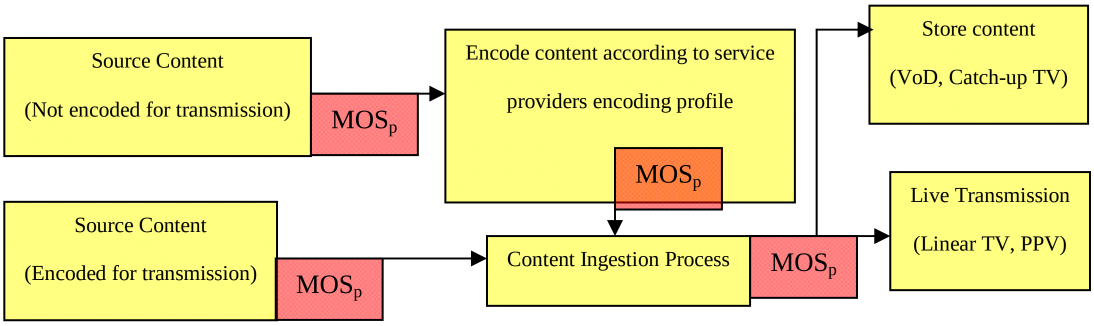


Figure 3: Post-transmission quality measurement

In addition to the three measurement stages identified above, a fourth form of measurement is commonly performed, namely during transmission planning. Parametric models have been defined to examine the impact of different transmission performance characteristics on the quality of speech (ITU-T Recommendation G.107 [i.42]) and videophone (ITU-T Recommendation G.1070 [i.12]) services. Work is in progress to define a general parametric model for video services.

Annex A: Bibliography

- ITU-T Recommendation P.862: "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs".
- ITU-T Recommendation P.862.2: "Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs".
- ITU-T Recommendation Y.1540: "Internet protocol data communication service - IP packet transfer and availability performance parameters".
- ITU-T Recommendation Y.1541: "Network performance objectives for IP-based services".
- ITU-R Recommendation BT.1683: "Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference".
- ITU-T Recommendation P.563: "Single-ended method for objective speech quality assessment in narrow-band telephony applications".
- IETF RFC 793: "Transport Control Protocol (TCP)".
- IETF RFC 768: "User Datagram Protocol (UDP)".
- IETF RFC 2616: "Hypertext Transfer Protocol -- HTTP/1.1".
- IETF RFC 2326: "Real-Time Streaming Protocol (RTSP)".
- IETF RFC 3550: "RTP: A Transport Protocol for Real-Time Applications".
- ETSI TR 102 493: "Speech and multimedia Transmission Quality (STQ); Guidelines for the use of Video Quality Algorithms for Mobile Applications".
- ETSI EG 202 765-3: "Speech and multimedia Transmission Quality (STQ); QoS and network performance metrics and measurement methods; Part 3: Network performance metrics and measurement methods in IP networks".

History

Document history		
V1.1.1	August 2011	Publication